

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:29:55 ; Search time 41 seconds  
(without alignments)  
1686.758 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELASAGVRGAA.....VFLPKRISKRQRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2731	100.0	519	20	AAV41114
2	2504.5	91.7	437	20	AAV41116
3	2317.5	84.9	537	20	AAV41127
4	2262.5	82.8	522	20	AAV41124
5	2255.5	82.6	522	20	AAV41123
6	1604.5	58.8	667	20	AAV41118
7	1592	58.3	563	20	AAV41125
8	1532	56.1	605	22	AAV41125
9	1523	55.8	594	22	AAV41125
10	1193.5	43.7	450	22	AAV41125

11	1144	41.9	533	20	AAV41128	Ricinus communis s
12	1134	41.5	494	20	AAV41119	Soybean sucrose tr
13	1126	41.2	501	20	AAV41126	Daucus carota sucro
14	1123.5	41.1	521	23	AAE15581	Alternative versio
15	1122	40.9	510	23	AAV41129	A. thaliana SUT4 p
16	1117	40.9	510	23	ABV90930	Herbicidally activ
17	1110	40.6	523	20	AAV41129	Vicia faba sucrose
18	1107	40.5	491	23	ABV93369	Herbicidally activ
19	1106.5	40.5	500	22	AAV41128	L. esculentum SUT4
20	1106.5	40.5	500	22	AAV41128	S. tuberosum SUT4
21	1106.5	40.5	500	22	AAV41128	Herbicidally activ
22	1106.5	40.5	500	22	AAV41128	Herbicidally activ
23	1106.5	40.5	500	22	AAV41128	Herbicidally activ
24	1088	39.8	507	21	AAV41128	Arabidopsis thalia
25	1088	39.8	507	21	AAV41128	Arabidopsis thalia
26	1088	39.8	507	21	AAV41128	Arabidopsis thalia
27	1088	39.8	507	21	AAV41128	Arabidopsis thalia
28	1088	39.8	507	21	AAV41128	Arabidopsis thalia
29	1075	39.4	520	23	AAE15307	Glycine max sucro
30	1070	39.2	512	23	ABV91556	Herbicidally activ
31	1014	37.1	232	20	AAV41122	Wheat sucrose tran
32	848.5	31.1	400	20	AAV41117	Rice sucrose trans
33	677	24.8	322	20	AAV41120	Soybean sucrose tr
34	653.5	23.9	350	21	AAV41120	Arabidopsis thalia
35	634.5	23.2	326	21	AAV41120	Arabidopsis thalia
36	564.5	20.7	302	21	AAV41120	Arabidopsis thalia
37	393.5	14.4	599	22	ABV60709	Drosophila melanog
38	369	13.5	214	21	AAV41120	Arabidopsis thalia
39	351	12.9	133	20	AAV41115	Corn sucrose trans
40	350	12.8	190	21	AAV41115	Arabidopsis thalia
41	333	12.2	748	22	AAV41115	Human polyptide
42	303.5	11.1	553	19	AAV1869	Amino acid encoded
43	303.5	11.1	553	19	AAV1869	Prostate tumour sp
44	303.5	11.1	553	21	AAV28527	Protein encoded by
45	303.5	11.1	553	21	AAV28527	Human immunogenic
46	303.5	11.1	553	22	AAV69763	Human prostate CDN
47	303.5	11.1	553	22	AAV69763	Human prostate tum
48	303.5	11.1	553	22	AAV69763	Human prostate-spe
49	303.5	11.1	553	22	AAV69763	Human prostate-spe
50	303.5	11.1	553	22	AAV69763	Human P501S invent
51	303.5	11.1	553	22	AAV69763	Prostate tumour an
52	303.5	11.1	553	23	ABV77575	Human mast cell re
53	303.5	11.1	553	23	ABV77575	Prostate cancer-as
54	303.5	11.1	553	23	ABV77575	Human L1-12 protei
55	303.5	11.1	553	23	ABV77575	Human PROST 03. H
56	303.5	11.1	553	23	AAU10324	Human breast tumou
57	303.5	11.1	1079	22	AAV74800	Prostate tumour an
58	293.5	10.7	595	22	AAV74800	Alpha prepro-P501S
59	293.5	10.7	710	23	AAV74800	Thioredoxin-ubiqui
60	282	10.3	166	20	AAV41121	Vernonia sucrose t
61	280	10.3	166	21	AAV41121	Arabidopsis thalia
62	246	9.0	439	21	AAV74548	Neisseria meningit
63	246	9.0	431	20	AAV74548	Neisseria meningit
64	245	9.0	429	21	AAV74548	Neisseria meningit
65	245	9.0	451	20	AAV74548	Neisseria meningit
66	234.5	8.6	428	21	AAV74548	Neisseria gonorrh
67	227.5	8.3	530	23	AAV50662	Thioredoxin-ubiqui
68	217.5	8.0	371	22	AAV69875	Human prostate CDN
69	217.5	8.0	371	22	AAV69875	P553S splice varia
70	217.5	8.0	371	23	ABV95335	Human P553S splice
71	215.5	7.9	371	22	AAE01362	Human gene 11 enco
72	215.5	7.9	371	23	ABV64105	Human albumin fusi
73	195	7.1	77	23	ABV91473	Herbicidally activ
74	191.5	7.0	400	22	AAV69907	Human prostate pro
75	191.5	7.0	400	22	AAV69907	Rail2-P501S-E2 cons
76	191.5	7.0	400	23	ABV95367	Rail2-P501S-E2 cons
77	177.5	6.5	326	23	ABV77571	Human mast cell re
78	163.5	6.0	619	21	AAV40554	Human ORFX ORF318
79	162.5	6.0	430	22	AAV96747	Putative P. abyss
80	155.5	5.7	469	21	AAV58289	Lung cancer associ
81	153	5.6	70	22	AAV58289	Sucrose transporte
82	150	5.5	430	22	AAV58289	Human secreted pro
83	147.5	5.4	490	23	AAE22906	Human transporter



CC The invention provides nucleic acid sequences (AA23124-223135) encoding  
 CC sucrose transport proteins (AA41114-41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.

XX SQ Sequence 497 AA;

Query Match 91.7%; Score 2504.5; DB 20; Length 497;  
 Best Local Similarity 94.2%; Pred. No. 8.2e-249;  
 Matches 467; Conservative 18; Mismatches 10; Indels 1; Gaps 1;

QY 24 DHVAPISLGRLLAGWAGGVOYQWALQSLTTPYVOTLGLSHALTSFWMWLCGPIAGLVV 83  
 DB 3 DHVAPISLGRLLAGWAGGVOYQWALQSLTTPYVOTLGLSHALTSFWMWLCGPIAGLVV 62  
 QY 84 QPLVGLYSDRCTSRWRRRPPILGCMILICVAVIVGFSDDIGAAAGDTRKHCSTLYHGPR 143  
 DB 63 QPLVGLYSDRCTSRWRRRPPILGCMILICVAVIVGFSDDIGAAAGDTRKHCSTLYHGPR 122  
 QY 144 WHAAIVVVLGFWLLDFNSNTVOGPARAMADLCHDHGSPSAANSIFCSMMALGNILGYSSG 203  
 DB 123 WHAAIVVVLGFWLLDFNSNTVOGPARAMADLCHDHGSPSAANSIFCSMMALGNILGYSSG 182  
 QY 204 STNNHWHKWFPLKTSACCEACANLKGAFVAVVFLVLTTLTIFAKEVPYRANENLPTT 263  
 DB 183 STNNHWHKWFPLKTSACCEACANLKGAFVAVVFLVLTTLTIFAKEVPYRANENLPTT 241  
 QY 264 KAGEVETEPTGPLAVLKGKDLPPGMPVSVLLTVAITWLSWFFPILYDTDMGRIYHGD 323  
 DB 242 KAGEVETEPTGPLAVLKGKDLPPGMPVSVLLTVAITWLSWFFPILYDTDMGRIYHGD 301  
 QY 324 PKGSNAQISAFNEGVRVAGGLLNSVLGFSSEFLIPEMCKRKVPRVVVTSNFMVCVAM 383  
 DB 302 PKGSNAQISAFNEGVRVAGGLLNSVLGFSSEFLIPEMCKRKVPRVVVTSNFMVCVAM 361  
 QY 384 AATALISFWSLRDYGHVQDAITANASIKAVCLVLFALFGLVPLAILYVFPFAVTAQLAAT 443  
 DB 362 AATALISFWSLRDYGHVQDAITANASIKAVCLVLFALFGLVPLAILYVFPFAVTAQLAAT 421  
 QY 444 RGGGGGLCTGVNLISIVIPQVIAAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLL 503  
 DB 422 RGGGGGLCTGVNLISIVIPQVIAAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLL 481  
 QY 504 PKISKROFRAVSAGGH 519  
 DB 482 PKISKROFRAVSAGGH 497

RESULT 3  
 AA41127  
 AA41127 standard; protein; 537 AA.

XX AA41127;

XX 17-JAN-2000 (first entry)

DE Oryza sativa sucrose transport protein.

XX Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
 KW carbohydrate transport; plant cell.

XX Oryza sativa.

XX WO9953068-A2.

XX 21-OCT-1999.

XX 07-APR-1999; 99WO-US07562.

XX 09-APR-1998; 98US-00811148.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA Allen SM, Hitz WD, Rafalski JA;  
 PI WPI; 1999-620432/53.  
 XX  
 XX PT New sucrose transport proteins from plants, useful for controlling  
 PT carbohydrate transport and distribution in plant cells -  
 XX Example 3; Page 59-61; 64pp; English.  
 XX The invention provides nucleic acid sequences (AA23124-223135) encoding  
 CC sucrose transport proteins (AA41114-41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.

XX SQ Sequence 537 AA;

Query Match 84.9%; Score 2317.5; DB 20; Length 537;  
 Best Local Similarity 81.2%; Pred. No. 1.6e-229;  
 Matches 435; Conservative 39; Mismatches 43; Indels 19; Gaps 3;

QY 1 MARGD-----GGQLAELSAGVRG-----AAVVDHVPISLGRLLIAGWVAG 42  
 DB 1 MARSGAGGGGGGGGGLLELSVGVGGGARGGGGGEAAAVETAAPISLGRLLISGMVAG 60  
 QY 43 GVQYGMALQSLTTPYVOTLGLSHALTSFWMWLCGPIAGLVVQPLVGLYSDRCTSRWRRR 102  
 DB 61 GVQYGMALQSLTTPYVOTLGLSHALTSFWMWLCGPIAGLVVQPLVGLYSDRCTSRWRRR 120  
 QY 103 PFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSN 162  
 DB 121 PYILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSN 180  
 QY 163 TVQGPAMADLCHDHGSPSAANSIFCSMMALGNILGYSSGTTNNHWHKWFPLKTSACCE 222  
 DB 181 TVQGPAMADLCHDHGSPSAANSIFCSMMALGNILGYSSGTTNNHWHKWFPLKTSACCE 240  
 QY 223 ACANLKGAFVAVVFLVLTTLTIFAKEVPYRANENLPTTAKAGEVETEPTGPLAVLKG 282  
 DB 241 ACANLKGAFVAVVFLVLTTLTIFAKEVPYRANENLPTTAKAGEVETEPTGPLAVLKG 299  
 QY 283 FKDLPPGMPVSVLLTVAITWLSWFFPILYDTDMGRIYHGDPKGSNAQISAFNEGVRVGA 342  
 DB 300 FRNLTGMPVSVLLTVAITWLSWFFPILYDTDMGRIYHGDPKGSNAQISAFNEGVRVGA 359  
 QY 343 FGLLNSVLGFSSEFLIPEMCKRKVPRVVVTSNFMVCVAMAAATLISFWSLRDYGHVQ 402  
 DB 360 FGLLNSVLGFSSEFLIPEMCKRKVPRVVVTSNFMVCVAMAAATLISFWSLRDYGHVQ 419  
 QY 403 DAITANASIKAVCLVLFALFGLVPLAILYVFPFAVTAQLAATRGGGGLCTGVNLISIVIP 462  
 DB 420 KAITADKSIKAVCLVLFALFGLVPLAILYVFPFAVTAQLAATRGGGGLCTGVNLISIVIP 479  
 QY 463 QVITAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLLPKISKROFRAVSAGG 518  
 DB 480 QVITAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLLPKISKROFRAVSAGG 535

RESULT 4  
 AA41124  
 ID AA41124 standard; protein; 522 AA.

XX AA41124;

XX 17-JAN-2000 (first entry)

XX DE Wheat sucrose transport protein (clone wlm24.pk0015.g11).  
 XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
 KW carbohydrate transport; plant cell.  
 XX OS Triticum aestivum.  
 XX PN W09953068-A2.  
 XX PD 21-OCT-1999.  
 XX PF 07-APR-1999; 99WO-US07562.  
 XX PR 09-APR-1998; 98US-0081148.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PI Allen SM, Hitz WD, Rafalski JA;  
 XX DR WPI: 1999-620432/53.  
 XX DR N-PSDB; AA223134.  
 XX PT New sucrose transport proteins from plants, useful for controlling  
 PT carbohydrate transport and distribution in plant cells -  
 XX PS Claim 5; Page 54-55; 64pp; English.  
 XX CC The invention provides nucleic acid sequences (AA223124-223135) encoding  
 CC sucrose transport proteins (AA41114-Y41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.  
 XX SQ Sequence 522 AA;  
 Query Match 82.8%; Score 2262.5; DB 20; Length 522;  
 Best Local Similarity 80.5%; Pred. No. 7.2e-224;  
 Matches 424; Conservative 40; Mismatches 50; Indels 13; Gaps 3;  
 QY 1 MARGGGQALAEISAGVRAA-----VVDHVPISLGRLLILAGMVAGVQVQWALQL 52  
 DB 1 MARGGGNGEVELSVGGGAGAGADAPVD----ISLGRLLILAGMVAGVQVQWALQL 56  
 QY 53 SLLTPVQVTLGLSHALTSPFMWLCGPVGLVQVPLVGLYSDRCTSRWRRRPPILTCMLI 112  
 DB 57 SLLTPVQVTLGLSHALTSPFMWLCGPVGLVQVPLVGLYSDRCTSRWRRRPPILTCMLI 116  
 QY 113 CVAVIVWGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAM 172  
 DB 117 CVAVIVWGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAM 176  
 QY 173 ADLDCDHGSPSAANSIFCSWMAALNLTGYSSTNNHWFPEPLKTSACCEANLKGAF 232  
 DB 177 ADLSAQHGSPSAANSIFCSWMAALNLTGYSSTNNHWFPEPLRTRACCEANLKGAF 236  
 QY 233 VAVFLVLCVLTTLTFAKEVPRANENLPTTKAGGEVETPTGVLAVLKGFKDLPMPGS 292  
 DB 237 VAVFLVLCVLTTLTFAKEVPRANENLPTTKAGGEVETPTGVLAVLKGFKDLPMPGS 295  
 QY 293 VLLVTAITWLSWFFPILYDTHMGREIYHGDPKGSNAQISAFNEGVRCVAGFLLNSVIL 352  
 DB 296 VLLVTAITWLSWFFPILYDTHMGREIYHGDPKGTPDEANAPQAGVAGFLLNSVIL 355  
 QY 353 GFSSFLIEPMCKRGVPRVWVTSNEMVCVMAATALISFWSLRDHYGVQDAITANASTK 412  
 DB 356 GFSSFLIEPMCKRGVPRVWVTSNEMVCVMAATALISFWSLRDHYGVQDAITANASTK 415  
 QY 413 AVCLVLFAPLGLVPLAILYSVPPFAVTAQAATRGGGGLCTGVNLNITSIVIPQVIALGAGP 472

DB 416 IVSLALFAFLGLPLAILYSVTFVTAQAANRCGGQWLCTGVNLNIAIPAQVIALGAGP 475  
 QY 473 WDALFGKGNIPAFGVASAFALVGGVGVVFLLPKISKROFRAVSGGH 519  
 DB 476 WDELFGKGNIPAFGVASAFALVGGVGVVFLLPKISKROFRAVSGGH 522  
 RESULT 5  
 AAY41123  
 ID AAY41123 standard; protein; 522 AA.  
 XX AC AAY41123;  
 XX DT 17-JAN-2000 (first entry)  
 XX DE Wheat sucrose transport protein (clone wlm24.pk0015.g11).  
 XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
 KW carbohydrate transport; plant cell.  
 XX OS Triticum aestivum.  
 XX PN W09953068-A2.  
 XX PD 21-OCT-1999.  
 XX PF 07-APR-1999; 99WO-US07562.  
 XX PR 09-APR-1998; 98US-0081148.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PI Allen SM, Hitz WD, Rafalski JA;  
 XX DR WPI: 1999-620432/53.  
 XX DR N-PSDB; AA223133.  
 XX PT New sucrose transport proteins from plants, useful for controlling  
 PT carbohydrate transport and distribution in plant cells -  
 XX PS Claim 5; Page 51-53; 64pp; English.  
 XX CC The invention provides nucleic acid sequences (AA223124-223135) encoding  
 CC sucrose transport proteins (AA41114-Y41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.  
 XX SQ Sequence 522 AA;  
 Query Match 82.6%; Score 2255.5; DB 20; Length 522;  
 Best Local Similarity 80.3%; Pred. No. 3.8e-223;  
 Matches 421; Conservative 42; Mismatches 54; Indels 7; Gaps 3;  
 QY 1 MARGGGQALAEISAGVRC-----AAVVDHVPISLGRLLILAGMVAGVQVQWALQLSLL 55  
 DB 1 MARGGGNGEVELSVGGGAGAGGEGQPAVDISLGRLLILAGMVAGVQVQWALQLSLL 60  
 QY 56 TPYVOTLGLSHALTSPFMWLCGPVGLVQVPLVGLYSDRCTSRWRRRPPILTCMLICVA 115  
 DB 61 TPYVOTLGLSHALTSPFMWLCGPVGLVQVPLVGLYSDRCTSRWRRRPPILTCMLICVA 120  
 QY 116 VIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAMMADL 175  
 DB 121 VIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAMMADL 180  
 QY 176 CDHGPCSAANSIFCSWMAALNLTGYSSTNNHWFPEPLKTSACCEANLKGAFVAV 235

Db 181 SAOHGSAANSIFCSWMAIGNILYSSGSTNNHKKWFPFLRTRACCEACANLKGAFVLAV 240  
 QY 236 VFLVLCITVTLFAKEVYPYRANENLPTTKAGGEVETPTGPLAVLKGFKDLPCGMPVLL 295  
 Db 241 LVAFCLVITVIFAKEIPIKAIPLP-TKGNQVEVEPTGPLAVFKGNLPP-MPSVLL 298  
 QY 296 VTAITWLSWFFIYLDYDWMGREIYHGDPKGSNAQISAFNEGVRVGAFGLLNSVILGFS 355  
 Db 299 VTGLTWLSWFFIYLDYDWMGREIYHGDPKGTDEANAFQAGVRAGAFGLLNSVILGFS 358  
 QY 356 SFLEPCMKRVGPRVWVTSNFMVVCVMAATAALISFSLRDYHGYVQDAITANASIKAVC 415  
 Db 359 SFLEPCCKRLGPRVWVSSNLFVCLSNAAICIIISWATODLHYIQHAIATASKEIKIVS 418  
 QY 416 LVLEAFGLVPLAILYSVPFAVTAQLAATRGGGGLCTGVNLISIVIPQVITIALGAGPWA 475  
 Db 419 LALFAFGIPLAILYSVPFAVTAQLAATRGGGGLCTGVNLIAIVIPQVITIAVAGPWA 478  
 QY 476 LFGKGNIPAFGVSFAFALVGVGVVFLLPKISKRQFRAVSAGGH 519  
 Db 479 LFGKGNIPAFGVSFAFALVGVGVVFLLPKISKRQFRAVSAGGH 522

RESULT 6  
 AAY41118  
 ID AAY41118 standard; protein; 667 AA.  
 XX AC AAY41118;  
 XX DT 17-JAN-2000 (first entry)  
 XX DE Rice sucrose transport protein (clone rls6.pk0076.e2).  
 XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
 XX KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
 XX KW carbohydrate transport; plant cell.  
 XX OS Oryza sativa.  
 XX PN W09953068-A2.  
 XX PD 21-OCT-1999.  
 XX PF 07-APR-1999; 99WO-US07562.  
 XX PR 09-APR-1998; 98US-0081148.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PI Allen SM, Hitz WD, Rafalski JA;  
 XX DR WPI; 1999-620432/53.  
 XX DR N-PSDB; AA223128.  
 XX PT New sucrose transport proteins from plants, useful for controlling  
 XX PS carbohydrate transport and distribution in plant cells -  
 XX Claim 5; Page 43-45; 64pp; English.

CC The invention provides nucleic acid sequences (AA223124-223135) encoding  
 CC sucrose transport proteins (AAY41114-Y41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.

SQ Sequence 667 AA;

Query Match 58.88; Score 1604.5; DB 20; Length 667;  
 Best Local Similarity 55.94; Pred. No. 5.5e-156;  
 Matches 299; Conservative 84; Mismatches 101; Indels 51; Gaps 5;

QY 33 RLILAGVAGGVOYQWALQLSLTPYVOTLGLSHALTSMFLCGPIAGLVVQPLVGLYSD 92  
 Db 133 KLVLAQWVAGVQWALQLSLTPYVOTLGLSHALTSMFLCGPIAGLVVQPLVGLYSD 192  
 QY 93 RCTSRWRRRRPFIITGCMILICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIIVYL 152  
 Db 193 KCRSKYGRRRPFILAGLMICFAVTLIGFSADLGVILGDTTEHCSLYHGPRHAAIIVYL 252  
 QY 153 GFLLDSSNTVQGPARAMMADLCHHGPSAANSIFCSWMAIGNILYSSGSTNNHKKW 212  
 Db 253 GFWMLDLANNTVQGPARAMMADLCHHGPSAANSIFCSWMAIGNILYSSGSTNNHKKW 312  
 QY 213 PFLTASACCEACANLKGAFVLAVVFLVLCITVTLFAKEVYPYR----- 255  
 Db 313 PFLTRACCEACSNLKAFLVAVVFLFCMSVTLTYFAEEIPLPTDAORLSDSAPLLNGS 372  
 QY 256 -----ANENLPTTKAGGE-VETEPTGPLAY-----LKGFK 284  
 Db 373 RDDNNASNEPRNGALPNCHTGDGSNVVPANSNAEDSNRENVEVDFGPGAVLVNLTSMR 432  
 QY 285 DLPGMPSVLLVTAITWLSWFFIYLDYDWMGREIYHGDPKGSNAQISAFNEGVRVGAFG 344  
 Db 433 HLPFGMYSVLLVMALTWLSWFFIYLDYDWMGREIYHGDPNUGNUSERKAYDNGVREGAFG 492  
 QY 345 LLLNSVILGSSFLTEPMCRKVGPRVWVTSNFMVVCVMAATAALISFSLRDYHGYVQDA 404  
 Db 493 LLLNSVILGSGFLVDPLCLRLMGLARLVWAINFTVVICMLATAILSWISFDLYSKLHHI 552  
 QY 405 ITANASIKAVCLVLFALVGLVPLAILYSVPFAVTAQLAATRGGGGLCTGVNLISIVIPQV 464  
 Db 553 IGANKTVKNSALIVFSLGLPLSITYSVFSTVTAELTAGTGGGGLATGVNLIAIVPQI 612  
 QY 465 IIALGAGPWAALFGKGNIPAFGVSFAFALVGVGVVFLLPKISKRQFRAVSAGGH 519  
 Db 613 VVSLGAGPWAALFGGNNVPALASVFSLGAGVLAVLKLPLD-PNSYR--SAGEH 664

RESULT 7  
 AAY41125  
 ID AAY41125 standard; protein; 563 AA.  
 XX AC AAY41125;  
 XX DT 17-JAN-2000 (first entry)  
 XX DE Wheat sucrose transport protein (clone wlmk1.pk0002.e11).  
 XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
 XX KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
 XX KW carbohydrate transport; plant cell.  
 XX OS Triticum aestivum.  
 XX PN W09953068-A2.  
 XX PD 21-OCT-1999.  
 XX PF 07-APR-1999; 99WO-US07562.  
 XX PR 09-APR-1998; 98US-0081148.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PI Allen SM, Hitz WD, Rafalski JA;  
 XX DR WPI; 1999-620432/53.  
 XX DR N-PSDB; AA223135.  
 XX PT New sucrose transport proteins from plants, useful for controlling  
 XX PS carbohydrate transport and distribution in plant cells -  
 XX Claim 5; Page 56-57; 64pp; English.



Db 362 VLVNLTSLRHLPMHSHVLIYVMTLWLPWPFPLFDTDMGMEVYHGDPKGEADENVAY 421  
 Qy 335 NEGVRGAGFGLLSVILGSESSFLIEPMCKRGVPRVWVTSNFMVCVMAATALISWSL 394  
 Db 422 NQGVREGAFGLLSVILGSSFLIEPMCKWIGSRLVMAVSNFVFCMACTAIISVSI 481  
 Qy 395 RYHGYQDAITANASIKAVCLVLFALGVPLALYSVPEAVTAQLAATRGSGGGLCTGV 454  
 Db 482 SAHTGVQHVIGATKSTQIAALVVSLLGIPLAVTYSVPESITAEITADAGGGOLAIGV 541  
 Qy 455 LNISIVPQVIALAGPMDALFCKGNIPAGVASAFALGVGVVGVVFLLPKISKRQFRAV 514  
 Db 542 LNLAIVLPMQVSVSLGAGPMDALFSGGNIPAFVLASLAALAGIFAMLRPLNLS--NFK-- 598  
 Qy 515 SAGGH 519  
 Db 599 STGFH 603

RESULT 9  
 AAG80012  
 ID AAG80012 standard; Protein; 594 AA.  
 AC AAG80012;  
 XX  
 DT 17-JAN-2002 (first entry)  
 XX  
 DE A. thaliana SUT2 protein.  
 XX  
 KW SUT2; sugar concentration; sugar transport; transgenic plant; herbicide;  
 KW flowering time; fruiting time; germination rate; cold tolerance;  
 KW drought tolerance; sweetness; branching length; photosynthesis;  
 KW pesticide; sugar metabolism.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200173086-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-EP02148.  
 XX  
 PR 24-MAR-2000; 2000DE-1014672.  
 PR 11-OCT-2000; 2000DE-1050233.  
 XX  
 PA (PROM/) FROMMER W.  
 XX  
 PI Frommer W, Ward JM, Weise A, Barker L, Schulze W, Kuehn C;  
 XX  
 DR WPI; 2001-611639/70.  
 DR N-PSDB; AAI68575.  
 XX  
 PT Modifying sugar flow and concentration in plants, useful e.g. for  
 PT increasing content of sugars and oils, by altering activity of a low  
 PT affinity, high capacity sugar transporter -  
 XX  
 PS Claim 21b; Page 89-91; 102pp; German.  
 XX  
 ST This invention describes a novel method for modifying the sugar flow  
 CC and/or sugar concentration in plant tissue in which the activity of a  
 CC sugar transporter (I) with high transport capacity but low affinity is  
 CC modified. At least one plant cell is transformed with a vector that  
 CC includes a nucleotide sequence (II) expression of which modifies  
 CC transport activity of (I). The transformed cells are then regenerated.  
 CC The method is used to produce transgenic plants that have increased sugar  
 CC or oil contents in sink organs and harvested materials, altered flowering  
 CC and fruiting times, increased germination rate, better cold and drought  
 CC tolerances, increased sweetness, reduced branching length, and/or  
 CC increased rate of photosynthesis. Nucleic acid sequences that encode (I)  
 CC are also useful for identifying modulators, especially inhibitors of  
 CC sugar transport (potentially useful as herbicides and pesticides) or  
 CC interactors that modify sugar transport, as molecular markers in  
 CC hybridization programs and for modifying the affinity of a protein,

CC especially of the SUC/SUT family, for a substrate, especially sucrose.  
 CC Also the central cytoplasmic loop of SUT2 is useful for regulation of  
 CC signal transduction, especially in sugar metabolism. The method makes  
 CC possible targeted alterations in sugar transport. This sequence  
 CC represents the Arabidopsis thaliana SUT2 protein described in the method  
 CC of the invention.  
 XX  
 SQ Sequence 594 AA;  
 Query Match 55.8%; Score 1523; DB 22; Length 594;  
 Best Local Similarity 53.6%; Pred. No. 1.1e-147;  
 Matches 288; Conservative 82; Mismatches 117; Indels 50; Gaps 5;  
 Qy 30 SLGLILAGVAGVOYQWALQLSLLPYVOTLGLSHALTSMFLMCGIAGLVVQPLVGL 89  
 Db 59 SLVTLVLCVAVAGVQFGWALQLSLLPYIOTLGHAFSSFIWLCGPITGLVVPFVGI 118  
 Qy 90 YSDRCTSRWRRRPFILTCMCLICVAVIVVGFSSDIGAALGDTKREHSLYHGRMHAAIV 149  
 Db 119 WSDKCTSKYGRRRPFILVGSFMSIAVIIIGFSADIGYLLGDSKEHCSTFKGTRAAV 178  
 Qy 150 YVLGFLLDFSNVTVOGPARAMMADLCHDHGSPSANSIFCSWMAIGNILGYSSGSTNNWH 209  
 Db 179 FIIGFWLLDLANNTVQGPARRALLADLSGPDQRNTANAVFCLMMAIGNILGFSAGSGKW 238  
 Qy 210 KWPFPLKTSACCEACANLKGAFVAVVFLVLCVTLVTLFAKEVPYRANE----- 258  
 Db 239 EWFFFLTSRACCAACGNLKAFLAVVFLTICTLVTFYFAKEIFPSTSKNPKTRIODSAPLL 298  
 Qy 259 -----NLPTTKAGG-----EVETEPGFLAV---LKG 282  
 Db 299 DDLQSKGLEHSLKNGTANGIKYERVERDTEQFGENSEHEQDETYYVDGPGSVLVNLLTS 358  
 Qy 283 FKDLPPGNPSVLLVTAITWLSWFFLITDMDGMEIYHGDPKGSNAQISAFNEGVRYGA 342  
 Db 359 LRHLPPAMHSHVLIYVMTLWLSWFFFLITDMDGMEIYHGDPTGDSLHNEIYDQVREGA 418  
 Qy 343 FGLLNSVILGFSFLIEPMCKRGVPRVWVTSNFMVCVMAATALISFWSLRDYGTVQ 402  
 Db 419 LGLLNSVILGFSFLIEPMCKRGVPRVWVTSNFMVCVMAATALISFWSLRDYGTVQ 478  
 Qy 403 DAITANASIKAVCLVLFALGVPLALYSVPEAVTAQLAATRGSGGGLCTGVNLISVIP 462  
 Db 479 YIMRGNETTTRTAAVIVFALLGFPLAITVSPFSVTAETADSGGQGLAIGVLNLAIVIP 538  
 Qy 463 QVIITAGAGPMDALFCKGNIPAGVASAFALGVGVVFLLPKISKROFRAVSAGGH 519  
 Db 539 QMIVSLGAGPMDQLFGGGLNLPFVLASVAFAGVIALQRLPTLSS-SFK--STGFH 592

RESULT 10  
 AAG80013  
 ID AAG80013 standard; Protein; 450 AA.  
 XX  
 AC AAG80013;  
 XX  
 DT 17-JAN-2002 (first entry)  
 XX  
 DE L. esculentum SUT2 protein fragment.  
 XX  
 KW SUT2; sugar concentration; sugar transport; transgenic plant; herbicide;  
 KW flowering time; fruiting time; germination rate; cold tolerance;  
 KW drought tolerance; sweetness; branching length; photosynthesis;  
 KW pesticide; sugar metabolism; tomato.  
 XX  
 OS Lycopersicon esculentum.  
 FH  
 FT Key Location/Qualifiers  
 FT Protein. 1.450  
 FT /note= "partial N-terminal fragment"  
 XX  
 PN WO200173086-A2.  
 XX

104-OCT-2001.  
 26-FEB-2001; 2001WO-EP02148.  
 24-MAR-2000; 2000DE-1014672.  
 11-OCT-2000; 2000DE-1050233.  
 (FROM/) FROMMER W.  
 Frommer W, Ward JM, Weise A, Barker L, Schulze W, Kuehn C;  
 WPI; 2001-611639/70.  
 DR N-PSDB; AA168576.  
 XX  
 XX  
 Modifying sugar flow and concentration in plants, useful e.g. for  
 increasing content of sugars and oils, by altering activity of a low  
 affinity, high capacity sugar transporter  
 Claim 21b; Page 91-92; 102pp; German.  
 XX  
 This invention describes a novel method for modifying the sugar flow  
 and/or sugar concentration in plant tissue in which the activity of a  
 sugar transporter (I) with high transport capacity but low affinity is  
 modified. At least one plant cell is transformed with a vector that  
 includes a nucleotide sequence (II) expression of which modifies  
 transport activity of (I). The transformed cells are then regenerated.  
 The method is used to produce transgenic plants that have increased sugar  
 or oil contents in sink organs and harvested materials, altered flowering  
 and fruiting times, increased germination rate, better cold and drought  
 tolerances, increased sweetness, reduced branching length, and/or  
 increased rate of photosynthesis. Nucleic acid sequences that encode (I)  
 are also useful for identifying modulators, especially inhibitors of  
 sugar transport (potentially useful as herbicides and pesticides) or  
 interactors that modify sugar transport, as molecular markers in  
 hybridization programs and for modifying the affinity of a protein.  
 Especially of the SUC/SUT family, for a substrate, especially sucrose.  
 Also the central cytoplasmic loop of SUT2 is useful for regulation of  
 signal transduction, especially in sugar metabolism. The method makes  
 possible targeted alterations in sugar transport. This sequence  
 represents the Lycopersicon esculentum SUT2 protein described in the  
 method of the invention.  
 XX  
 XX  
 Sequence 450 AA;  
 Query Match 43.7%; Score 1193.5; DB 22; Length 450;  
 Best Local Similarity 51.7%; Pred. No. 6.5e-114;  
 Matches 233; Conservative 59; Mismatches 102; Indels 57; Gaps 5;  
 QY 123 SDIGALGDTKEHCYHGRHAAIVYVGLFWLDFSNVTQGPARAMADLCHDHGFS 182  
 DB 1 ADIGYLLGDTKEHCSTFKGTRRAAIVFVVGFWMLDLANTVQGPALLADLSGPDOR 60  
 QY 183 AANSIFCSWMAIGNILGYSGSTNNHKKFPFLKTSACCACANLKGAFVAVVFLVCL 242  
 DB 61 TANAVFCSMVAVGNILGFSAGSGGWHFPFLTRACCPCGNLKAALVAVVFLTCL 120  
 QY 243 TVTLIFAKEVPY-----RA 256  
 DB 121 LVTLIFANEVPLSPKQYKRLSDSAPLLDSDPONTGFDLSKRELQSVNSVANNESEMRV 180  
 QY 257 NENLPTT-----KAGVEVETETGPLAV--LKGFKDLPPGMPSVLLVTAITWLSWFPPT 308  
 DB 181 ADNSPKNEORPDKDQGSFADSPGAVLVNLLTSLRHLPPAMHSHVLIVMALWLPWFPEF 240  
 QY 309 LYDTDMGREIYHGDPKGSNAQISAFNEGVRCAGFLLNSYLFGSSFLIEPMCKVKCP 368  
 DB 241 LFTDWMGREVYHGDPKGDEAVNAYNQGVREGAFGLLNSVVLGSSFLIEPMCKWIGS 300  
 QY 369 RVVWVTSNFMVCMVAMATALISFWSURDYHGVYQDITANASTKAVCLVLFVAFGLVPLAI 428  
 DB 301 RLWAVSNFIVFCMACTAISVWSISATEGVQHVIGATKSTQIAALVVSLLGIPLAV 360  
 QY 429 LYSVPFAVTAQLAATRGGGGLCTGVLNTSIVIPQVIAALGAGPMDALFGKGNIPAGVA 488

DB 361 TYSVPFSITAEITADAGGGQGLAIGVLNLAIVLPQVWVSLGAGPMDALFGGGINPAFVLA 420  
 QY 489 SAFALVGGVGVFLLPKISKROFRAVSAGGH 519  
 DB 421 SLAALAAGIFAMLRUPNIUSS-NFX--STGCFH 448  
 RESULT 11  
 AAY41128  
 ID AAY41128 standard; protein; 533 AA.  
 XX  
 AC AAY41128;  
 XX  
 XX 17-JAN-2000 (first entry)  
 XX Ricinus communis sucrose transport protein.  
 XX  
 KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
 KW carbohydrate transport; plant cell.  
 OS Ricinus communis.  
 XX  
 PN WO9953068-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-US07562.  
 XX  
 PR 09-APR-1998; 98US-0081148.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA  
 PI Allen SM, Hitz WD, Rafalski JA;  
 XX  
 DR WPI; 1999-620432/53.  
 XX  
 PT New sucrose transport proteins from plants, useful for controlling  
 PT carbohydrate transport and distribution in plant cells -  
 XX  
 PS Example 3; Page 61-62; 64pp; English.  
 XX  
 CC The invention provides nucleic acid sequences (AA23124-23135) encoding  
 CC sucrose transport proteins (AAY4114-Y41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.  
 XX  
 SQ Sequence 533 AA;  
 Query Match 41.9%; Score 1144; DB 20; Length 533;  
 Best Local Similarity 45.0%; Pred. No. 1e-108;  
 Matches 226; Conservative 99; Mismatches 151; Indels 26; Gaps 8;  
 QY 19 AAVVDHVAIPISLGRLLLAGWAGVOYGVWALQSLTPYVOTLGLSHALTSFMLCGPI 78  
 DB 25 AGAAEPNSSP--LRKVVWVASAAGIQFGWALQSLTPYVOLLGIPHTWAAFIWLCGPI 82  
 QY 79 AGLVQPLVGLYSDRCTSRWRRRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSL 138  
 DB 83 SGMVQPIVGYHSDRCTSRFRRRPFIAAGAAFAVLAFLICYAADLGLHSGD----SL 137  
 QY 139 YHGRWHAIVVVLGFWLDFSNVTQGPARAMADLCHDHGFS-----AANSIFCSWMA 193  
 DB 138 DKSPKTRAIATFVWGFILDVANMLQGPCRALLADL---SGTSOKTKRTANALSFEMA 194  
 QY 194 LGNILGYSGSTNNHKKFPFLKTSACCACANLKGAFVAVVFLVCLTVTLIFAKEVP 253  
 DB 195 VGNVLGYAAGATHYLKLFPFTKTTACDVCYCANLKSCFFISIVLILLTLVLTALSVKEP 254





CC sucrose transport proteins (AAV41114-Y41125) derived from corn, rice,  
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
 CC be recombinantly expressed by standard recombinant methodology. The  
 CC invention facilitates studies on carbohydrate metabolism and function in  
 CC plants, provides genetic tools for the manipulation of these  
 CC biosynthetic pathways, and provides a means to control carbohydrate  
 CC transport and distribution in plant cells.

XX Sequence 501 AA;

Query Match 41.2%; Score 1126; DB 20; Length 501;  
 Best Local Similarity 44.6%; Pred. No. 6.7e-107;  
 Matches 225; Conservative 102; Mismatches 142; Indels 36; Gaps 8;

QY 18 GAAVVDHVAITSLRLIAGVAGVOYGMALQSLTLPYVQTLGLSHALTSPFWLQGP 77  
 DB 14 GATAAPPPRSVLSRLRLRVASVACIQGQWALQSLTLPYVQELGHPHAWSSIIWLCGP 73  
 QY 78 IAGLVVQPLVGLYSRDRCTSRGRRRPFILTGCMILICVAVIVVGFSSDGAALGDTKEHCS 137  
 DB 74 LSGLLVQPIVGHMSDQCTSKYGRRRPFIVAGGTATILAVIIIAHSADIGLLGDTADNKT 133  
 QY 138 LYHGPWRHAAIVVGLFWLLDSNTVQGPARAMWADLC--DHHGPSAANSIFCSWMAIG 195  
 DB 134 M-----AIVAFVIGFWILDVANNTQGPCRALLADLTGNDARRTRVANAYESLFMAIG 186  
 QY 196 NILGYSSGSTNNHWFPLKTSACCEACANLKGAFVAVFLVCLTVTLIFAKEVPYR 255  
 DB 187 NVLGATGAYSGWKVPFSLTSSCTINCANKSAFYDIIIFIIITYSISAAKERPI 246  
 QY 256 ANENLPTTKAGGEVET---EPTGPLAVLKGFKDLPMPSPVLLVTAITWLSWFFILYDT 312  
 DB 247 SSODGQFSEDGTAQSGHIEAFWLFCFTLRLLPGSVVILLVTCNLWIGWFFILFDT 306  
 QY 313 DWNGRIYHGDGKGSNAQISAFNEGVRVCAFGLLNSVILGFSSFLIEPMCRKVGPRVW 372  
 DB 307 DWNGRIYHGEPE---QOQSYSDGVGMGAFGLMNSVVLGITSVLMEKLRWISGGMW 362  
 QY 373 VTSNFVVCVMAATAISFWSLR-DY-----HGVYQDAITANASTKAVCLVLFALGVP 425  
 DB 363 GLSNILMTICFFAMLLITFIANKMDYGTNPPNGIVISA-----LIVEALIGP 411  
 QY 426 LAILYVPRVAVTAQLAATRGGOGLCTGVNLNISTIVIPQVIAIAGAGPMDALFGKNIPAF 485  
 DB 412 LAITYSPVALYSTRIESLGLQGLSMGLNLAIIVPQIVISLGSQPMWDLQGGNSPAP 471  
 QY 486 GVASAFALVGGVGVFLL--PKISK 508  
 DB 472 VVAALSAFAGLIALAIRPRVDR 496

RESULT 14  
 AAEL15581  
 ID AAEL15581 standard; Protein; 521 AA.

XX

AC AAEL15581;

DT 12-MAR-2002 (first entry)

DE Alternative version of GmsUT1 protein.

XX Glycine max ankyrin-related protein; GMA; GmsUT1; sugar transport;

KW sucrose/proton symporter; soybean ankyrin-related protein; SAR; ANK;

XX sugar allocation; nutritional value.

OS Glycine max.

XX Key Location/Qualifiers

FT Region 229..279

XX /note= "Cytosolic loop"

PN WO2001188139-A2.

PD 22-NOV-2001.  
 XX  
 PF 11-MAY-2001; 2001WO-US15315.  
 XX  
 PR 12-MAY-2000; 2000US-203974P.  
 XX  
 PA (UNITW ) UNIV WASHINGTON STATE RES FOUND.  
 XX  
 PI Grimes HD, Elmer AM, Murphy KA;  
 XX  
 DR WPI; 2002-062385/08.  
 XX  
 DR N-PSDB; AAD24773.  
 XX  
 PT New purified protein having Glycine max ankyrin-related (GMA) protein  
 PT biological activity, useful to alter GMA levels in plants to confer  
 PT altered sugar transport and/or altered sugar allocation properties  
 XX  
 PS Claim 13; Page 51-53; 60pp; English.  
 XX  
 CC The invention relates to (soybean) Glycine max (Gm) ankyrin (ANK)-  
 CC related proteins 1 and 2 referred to as GMA1 and GMA2 and nucleic acid  
 CC molecules encoding them. GMA proteins also known as soybean ankyrin-  
 CC related (SAR) proteins interact with Gm sucrose/H+ (proton) symporter  
 CC designated as (SUT1). Manipulating the expression of GmsUT1 and GMA  
 CC in plants is useful to confer altered sugar transport and/or altered  
 CC sugar allocation properties. Alteration of GMA protein levels in plants  
 CC could be used to increase the nutritional value of plant tissues, for  
 CC instance plant seeds or grain. The present sequence is an alternative  
 CC version of GmsUT1 protein.  
 CC Note: This sequence is stated to be the same as that shown as SEQ ID NO:2  
 CC (AAE15307) in Fig 1B of the specification. However the sequences differ  
 CC at several positions.  
 XX  
 SQ Sequence 521 AA;

Query Match 41.1%; Score 1123.5; DB 23; Length 521;  
 Best Local Similarity 44.0%; Pred. No. 1.3e-106;  
 Matches 221; Conservative 104; Mismatches 148; Indels 29; Gaps 9;

QY 31 LGRLLIAGVAGVOYGMALQSLTLPYVQTLGLSHALTSPFWLQGPVQLVGLY 90  
 DB 31 LRKIMVVASIAGVOYGMALQSLTLPYVQTLGLSHALTSPFWLQGPVQLVGLY 90

QY 91 SDRCTSRGRRRPFILTGCMILICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIYV 150  
 DB 91 SDRCTSRGRRRPFILTGCMILICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIYV 150

QY 151 VLGFWLLDSNTVQGPARAMWADLC--DHHGPSAANSIFCSWMAIGNLGYSSTNNW 208  
 DB 146 VVGFWLLDSNTVQGPARAMWADLC--DHHGPSAANSIFCSWMAIGNLGYSSTNNW 208

QY 209 HKWFFPLKTSACCEACANLKGAFVAVFLVCLTVTLIFAKEVPYRANENLPTTKAGE 268  
 DB 206 HNVFFTKKADYVCANLKSCTFLSIALLTLSTIALTYVKEKTVSEK---TVRSVVE 262

QY 269 VETEPTG-PL--AVLKGFKDLPMPSPVLLVTAITWLSWFFILYDTWDMGREIYHGDPK 325  
 DB 263 EDGSHGMPGCGQLFGAFRELKRPWMLLVLTCLNIAFWFPELLFDTWDMGREVY---G 318

QY 326 GSAQISAFNEGVRVCAFGLLNSVILGFSSFLIEPMCRKV--GPRVWVTSNFWVCVAMA 384  
 DB 319 GTVGEKAYDRVRAGALGLMLNSVILGFSSFLIEPMCRKV--GPRVWVTSNFWVCVAMA 384

QY 385 ATALIS-----FWSLRDHYGVYQDAITANASTKAVCLVLFALGVPVPLATLYSVPRAV 436  
 DB 379 MTVLVTKMAQHPRQYTLPLNAH---OEPLPPPAAKAGALFLSLIGLPLATLYSIPAL 435

QY 437 TQAALATRGGOGLCTGVNLNISTIVIPQVIAIAGAGPMDALFGKNIPAFVASAFALVGG 496  
 DB 436 ASIFSSTSGAGGLSLGVLNLAIIVPQIVISLGSQPMWDLQGGNSLPAFVGVCAVAASG 495

QY 497 VVGFVLLPKISKRPRAVSAGG 518  
 DB 497 VVGFVLLPKISKRPRAVSAGG 518



CC useful as herbicides.

XX Sequence 510 AA;

Query Match 40.9%; Score 1117; DB 23; Length 510;

Best Local Similarity 47.4%; Pred. No. 5.8e-106;

Matches 225; Conservative 85; Mismatches 145; Indels 20; Gaps 8;

QY 34 LILAGVAGVGVGVALQSLTTPYVOTGLSHALTSEFMWLCGPIAGLVVQPLVGLYSR 93

DY 43 LRVASVAGGIQFGVALQSLTTPYVOTGLSHALTSEFMWLCGPIAGLVVQPLVGLYSR 102

QY 94 CTSWRGRRPFIITGCMILICVAVIVVGFSSDICAALGDKHCPSLYHGPRWHAIVVVLG 153

DY 103 CTSKYGRRRPFIVAGAVATISVMVIGHAADIGWAFGDREGKIK---PR--AIVAFVLG 156

QY 154 FWLLDSNNTVGVGPARAMADLC--DHHGPSAANSIFCSWMLGNILGYSSGSTNNWHKW 211

DY 157 FWLLDVANNMTQPCRALDLADLTENDNRTRVANGYFSLFMAVGNVLYGATGSYNGWYKI 216

QY 212 FPELKTSAACEACANLKGAFLVAVVFLVCLTVTLIFAKEVPRANENLPTTKAGEVET 271

DY 217 FTFKTVACNVECANLKSAFYIDVFIATITLSVAAHEVPLAS----LASEAHGQTS 272

QY 272 EPTGPLAVLKG-FKOLPPGMPSVLLTATITWLSWPFILYDITDMGREGIYHGDPKGSNAQ 330

DY 273 TDEAFLEIFGTFRYPFGNWIILLTALTWTIGWPFILFDITDMGREGIYGPENIG--- 329

QY 331 ISAFNEGVRGAGLLNSVILGFSFSLIEPCKRKVGPRVWVTSNFMVYCVAMAATALLS 390

DY 330 -TSYAGVSGMALGLMLNSVFLGITSVLMKLCRWGAFVNGISNLMALICFLGMIITS 388

QY 391 FWSLRDHYGVODAITANASIKAVCLVLFALGVPLAILYSVPFAVTAQLAATRGGGQGL 450

DY 389 F--VASHLGYIGHE-QPPASIVFAAVLIETILGIPLAITYSVPYALISRIESLGLQGL 445

QY 451 CTGVLNISVIVPOTIALGAGPMDALFGKGNIPAFGVASAFALVGVGVFLLPK 505

DY 446 SLGVLNLAIVIPQIVSVSGWDQLFGGNSPALAVGAATFGIGIVAILALPR 500

RESULT 17

AA411129

ID AAY41129 standard; protein; 523 AA.

XX AC

XX AAY41129;

XX 17-JAN-2000 (first entry)

XX DE Vicia faba sucrose transport protein.

XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;

XX KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;

XX KW carbohydrate transport; plant cell.

XX OS Vicia faba.

XX PN W09953068-A2.

XX PD 21-OCT-1999.

XX PF 07-APR-1999; 99WO-US07562.

XX PR 09-APR-1998; 98US-0081148.

XX PA (DUPO ) DU PONT DE NEMOURS &amp; CO E I.

XX PI Allen SM, Hitz WD, Rafalski JA;

XX DR WPI; 1999-620432/53.

XX PT New sucrose transport proteins from plants, useful for controlling

XX carbohydrate transport and distribution in plant cells -

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Example 3; Page 63-64; 64pp; English.

The invention provides nucleic acid sequences (AA23124-223135) encoding sucrose transport proteins (AA41114-Y41125) derived from corn, rice, soybean, Vernonia and wheat tissues. The sucrose transport proteins can be recombinantly expressed by standard recombinant methodology. The invention facilitates studies on carbohydrate metabolism and function in plants, provides genetic tools for the manipulation of these biosynthetic pathways, and provides a means to control carbohydrate transport and distribution in plant cells.

SQ Sequence 523 AA;

Query Match 40.6%; Score 1110; DB 20; Length 523;

Best Local Similarity 44.2%; Pred. No. 3.1e-105;

Matches 221; Conservative 104; Mismatches 155; Indels 20; Gaps 8;

QY 28 PISLGRILAGVAGVGVGVALQSLTTPYVOTGLSHALTSEFMWLCGPIAGLVVQPLV 87

DY 31 PSLPKIMVVASIAAGVQFGVALQSLTTPYVOTGLSHALTSEFMWLCGPIAGLVVQPLV 90

QY 88 GLYSRDRSRRGRRPFIITGCMILICVAVIVVGFSSDICAALGDKHCPSLYHGPRWHA 147

DY 91 GYHSDRSTRFRGRRPFIITGCMILICVAVIVVGFSSDICAALGDKHCPSLYHGPRWHA 145

QY 148 IYVVLGFWLLDSNNTVGVGPARAMADLC--DHHGPSAANSIFCSWMLGNILGYSSGST 205

DY 146 GFVVGFWLLDVANNMTQPCRALDLADLTENDNRTRVANGYFSLFMAVGNVLYGATGS 205

QY 206 NNHMKWFPPLKTSACEACANLKGAFLVAVVFLVCLTVTLIFAKEVPRANENLPTTKA 265

DY 206 SKLYHVFPFTKACNVYCANLKSCFFLSIALTLVLTATISALTYVKTALTPKTVTTED 265

QY 266 GGEVETET-GPLAVLKGFKDLPPGMPSVLLTATITWLSWPFILYDITDMGREGIYHGD 324

DY 266 GSSGGMPCFGOLS--GAFKELKRPMMIILLVTLCLNIAWFFLLFDITDMGKEVY--- 319

QY 325 KGSNAQISAFNEGVRGAGLLNSVILGFSFSLIEPCKRKV-GPRVWVTSNFMVYCVAM 383

DY 320 GGTGEGHAYDMGVREGALGLMLNSVILGATSLGVDILARGVGVKRLWGINVFLAICL 379

QY 384 AATALLISFWS--LRDY---HGVQDAITANASIKAVCLVLFALGVPLAILYSVPFAVTA 438

DY 380 GLTVLTVTLKLAQHSROYAPGTGALGDLPESEGIKAGALTFLSVLGVPLAITYSIPFALAS 439

QY 439 QLAATRGSGGLCTGVNLISVIVPOTIALGAGPMDALFGKGNIPAFGVASAFALVGVV 498

DY 440 IFSSTSGAGGGLSLGVNLAIVIPQIVSVSGWDQLFGGNSPALAVGAATFGIGIVAILALPR 499

QY 499 GVFLLPKISKROFRAVSAGG 518

DY 500 SIILLPSPPPDMAKSVSATG 519

RESULT 18

ABB93369

ID ABB93369 standard; Protein; 491 AA.

XX AC

XX ABB93369;

XX DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 2580.

XX KW Herbicidal; plant; agriculture; herbicide.

XX OS Arabidopsis thaliana.

XX PN WO200210210-A2.

XX PD 07-FEB-2002.

XX











PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
PR 15-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143442.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0132487.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.



PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
Query Match 39.8%; Score 1088; DB 21; Length 512;					
Best Local Similarity 43.6%; Pred. No. 5.6e-103;					
Matches 220; Conservative 100; Mismatches 159; Indels 26; Gaps 10;					
Qy	12	LSAGVGAADVHVA-----PISLGRLLTAGVAGVOYGNALQSLSLTTPVQTLGSHA 67			
Db	6	MEKAANGASALETQTGELDQPERLKRKIISVSSIAAGVQFGNALQSLSLTTPVQLLGIPIK 65			
Qy	68	LTSEFMWLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFLITGCMILICVAVIVVGFSSDIGA 127			
Db	66	WASLIWLCGPISGMLVQPIVGYHSDRCTSRFGRRRPFIIVAGAGLVTVAVFLIGYAADIGH 125			
Qy	128	ALGDTKEHCSLYHGPRHAAIVYVGLFWLLDFSNNTVOGPARAMMADLCHDHGPPS--AAN 185			
Db	126	SMGD-----QLDKPKPTRAIFAIFALGFWDVANNLTGGPCRAFADIASAGNAKTRTAN 180			
Qy	186	SIFCSWMLGNILYSSGSTNNHKKFPFLKTSACCEACANKLGAFLVAVFLVLCITVT 245			
Db	181	AFFEFFMAVGNVLGYAAGSYRNLYKVPFTWTESCDLYCANLKTCFFISITLLILVTFVS 240			
Qy	246	LIFAKEVPYRANENLPTTKAGGEVETEPTGPLAVILKGFKLPPGMPSPVLLVTAITWLSWF 305			
Db	241	LCYVKEKPW-----TPEPTADGKASNVPFFG-EIFGAFKELKRPMMMLIVTALNIAWF 294			
Qy	306	PFILYDTHMGREIYHG--DPKGSNAQISAFNEGVRVCAFGLLLSVILGFSSFLIEPMC 363			
Db	295	PFLLEDTHMGREYVGGNSDATATAASKKLYNDGVRAGALGLMLNAIYVGLFMSLGVENIG 354			
Qy	364	RKV-GPRVVVWTSNFMVCVANAATALISFWS--LRDHYGVYODAITANASTKACVCLVF 419			
Db	355	RKLGAKRLWGIIVFILAICIAMTVVTKQAEHNRD-HGGAKTGPPGN--VTAGALTLF 411			
Qy	420	AFLGVPLAILYVSPFAVTAQLAATRGQGGLCTGVNLNISIVIPQVITIALGAGPMDALFGK 479			

Db 412 AILGIPQAITFSIPFALASIFSTNSGAGLSGLVNLAIWVPMVISVGGPFDEFGG 471  
QY 480 GNIPAFGVASAFALYGVGVFLLP 504  
Db 472 GNIPAFVLGAIAAAVSGVIALTVLP 496

Search completed: April 17, 2003, 11:31:22  
Job time : 46 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:30:00 ; Search time 17 Seconds  
(without alignments)  
898.265 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGGGQLAELSGVRGAA.....VELLPKISKROFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_5/ptodata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_5/ptodata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_5/ptodata/1/iaa/6A\_COMB.pep.\*

4: /cgn2\_5/ptodata/1/iaa/6B\_COMB.pep.\*

5: /cgn2\_5/ptodata/1/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_5/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1079	39.5	516	1	US-08-356-340-4
2	1079	39.5	516	2	US-08-786-555-4
3	1075	39.4	525	1	US-08-356-340-2
4	1075	39.4	525	2	US-08-786-555-2
5	303.5	11.1	553	4	US-09-020-956-113
6	303.5	11.1	553	4	US-09-030-607-113
7	303.5	11.1	553	4	US-09-605-785-113
8	303.5	11.1	553	4	US-09-439-313-113
9	303.5	11.1	553	4	US-09-352-616A-113
10	303.5	11.1	553	4	US-09-602-877A-101
11	303.5	11.1	553	4	US-09-232-149A-113
12	217.5	8.0	371	4	US-09-605-785-708
13	126.5	4.6	618	4	US-08-595-553A-2
14	114	4.2	435	6	5268463-9
15	112	4.1	467	4	US-09-134-001C-3020
16	112	4.1	528	2	US-08-403-852B-21
17	112	4.1	528	3	US-08-510-646B-22
18	112	4.1	528	4	US-09-231-818-21
19	106	3.9	457	2	US-08-882-704A-6
20	106	3.9	457	4	US-09-151-957-6
21	105.5	3.9	436	6	5432081-10
22	103.5	3.8	474	4	US-09-332-041-5
23	101	3.7	563	2	US-09-031-392-2
24	101	3.7	563	4	US-09-299-549-2
25	101	3.7	563	4	US-09-610-417-2
26	100	3.7	323	4	US-09-134-001C-4635
27	100	3.7	457	6	5268463-7

28	100	3.7	486	4	US-09-291-922-10	Sequence 10, Appl
29	100	3.7	620	1	US-08-301-722A-2	Sequence 2, Appl
30	99.5	3.6	456	6	5432081-7	Patent No. 5432081
31	99	3.6	255	4	US-09-071-710-36	Sequence 36, Appl
32	99	3.6	255	4	US-09-525-397-36	Sequence 36, Appl
33	99	3.6	619	1	US-07-762-132A-2	Sequence 2, Appl
34	99	3.6	619	1	US-08-301-722A-4	Sequence 4, Appl
35	97	3.6	326	4	US-09-134-001C-4837	Sequence 4837, Ap
36	97	3.6	411	4	US-09-134-001C-3299	Sequence 3299, Ap
37	97	3.6	426	6	5268463-8	Patent No. 5268463
38	96	3.6	539	4	US-09-291-922-26	Sequence 26, Appl
39	95.5	3.5	434	2	US-08-677-049-7	Sequence 7, Appl
40	95	3.5	363	1	US-08-148-209A-2	Sequence 2, Appl
41	95	3.5	635	1	US-07-879-617A-9	Sequence 9, Appl
42	95	3.5	635	1	US-08-753-985-9	Sequence 9, Appl
43	94.5	3.5	353	4	US-09-077-675A-3	Sequence 3, Appl
44	93.5	3.4	428	6	5432081-9	Patent No. 5432081
45	93.5	3.4	524	4	US-09-134-001C-5457	Sequence 5457, Ap
46	92.5	3.4	347	4	US-09-134-001C-3960	Sequence 3960, Ap
47	92	3.4	363	1	US-08-148-209A-3	Sequence 3, Appl
48	91.5	3.4	316	1	US-08-118-270-46	Sequence 46, Appl
49	91.5	3.4	316	5	PCT-US93-08528-46	Sequence 46, Appl
50	91.5	3.4	437	4	US-09-134-001C-3907	Sequence 3907, Ap
51	91.5	3.4	599	1	US-08-301-722A-5	Sequence 5, Appl
52	90.5	3.3	599	1	US-08-295-814E-11	Sequence 11, Appl
53	90.5	3.3	599	3	US-08-240-783B-4	Sequence 4, Appl
54	90.5	3.3	599	3	US-09-084-813-4	Sequence 4, Appl
55	90.5	3.3	599	4	US-09-343-361-11	Sequence 11, Appl
56	90.5	3.3	599	5	PCT-US92-09662-4	Sequence 4, Appl
57	90	3.3	616	1	US-07-879-617A-13	Sequence 13, Appl
58	90	3.3	616	4	US-08-753-985-13	Sequence 13, Appl
59	89.5	3.3	364	4	US-09-077-675A-16	Sequence 16, Appl
60	89	3.3	667	1	US-07-879-617A-8	Sequence 8, Appl
61	89	3.3	667	1	US-08-753-985-8	Sequence 8, Appl
62	88.5	3.2	388	1	US-08-087-772A-2	Sequence 2, Appl
63	88.5	3.2	400	1	US-08-351-473B-5	Sequence 5, Appl
64	88.5	3.2	400	4	US-08-450-962-4	Sequence 4, Appl
65	88.5	3.2	493	1	US-08-362-512A-4	Sequence 4, Appl
66	88.5	3.2	493	4	US-08-964-939-4	Sequence 4, Appl
67	88.5	3.2	493	4	US-09-134-001C-5101	Sequence 5101, Ap
68	88	3.2	415	4	US-09-030-267-5	Sequence 5, Appl
69	88	3.2	418	4	US-08-996-644-2	Sequence 2, Appl
70	88	3.2	1308	2	US-09-352-552-2	Sequence 2, Appl
71	88	3.2	1308	3	US-09-352-552-2	Sequence 2, Appl
72	87.5	3.2	298	1	US-08-118-270-76	Sequence 76, Appl
73	87.5	3.2	298	5	PCT-US93-08528-76	Sequence 76, Appl
74	87.5	3.2	500	2	US-09-031-392-7	Sequence 7, Appl
75	87.5	3.2	500	4	US-09-299-549-7	Sequence 7, Appl
76	87.5	3.2	500	4	US-09-610-417-7	Sequence 7, Appl
77	87.5	3.2	627	1	US-08-295-814E-4	Sequence 4, Appl
78	87.5	3.2	627	1	US-08-291-299-10	Sequence 10, Appl
79	87.5	3.2	627	4	US-09-343-361-4	Sequence 4, Appl
80	87.5	3.2	627	5	PCT-US93-01959-4	Sequence 4, Appl
81	87.5	3.2	627	5	PCT-US95-10579-10	Sequence 10, Appl
82	87.5	3.2	632	1	US-08-295-814E-10	Sequence 10, Appl
83	87.5	3.2	632	1	US-09-343-361-10	Sequence 10, Appl
84	87.5	3.2	632	3	PCT-US93-01959-10	Sequence 10, Appl
85	87.5	3.2	4551	3	US-09-320-878-1	Sequence 1, Appl
86	87.5	3.2	4613	4	US-09-105-537-31	Sequence 31, Appl
87	87.5	3.2	11877	4	US-09-105-537-6	Sequence 6, Appl
88	87	3.2	878	4	US-09-463-238-4	Sequence 4, Appl
89	86.5	3.2	473	4	US-09-134-001C-3564	Sequence 3564, Ap
90	86.5	3.2	473	4	US-09-134-001C-3487	Sequence 3487, Ap
91	86.5	3.2	509	2	US-09-031-392-6	Sequence 6, Appl
92	86.5	3.2	509	4	US-09-299-549-6	Sequence 6, Appl
93	86.5	3.2	509	4	US-09-610-417-6	Sequence 6, Appl
94	86.5	3.2	553	3	US-08-501-572-3	Sequence 3, Appl
95	86.5	3.2	553	3	US-09-040-444-3	Sequence 3, Appl
96	86.5	3.2	599	1	US-07-879-617A-10	Sequence 10, Appl
97	86.5	3.2	599	1	US-08-753-985-10	Sequence 10, Appl
98	86.5	3.2	614	1	US-08-295-814E-12	Sequence 12, Appl
99	86.5	3.2	614	4	US-09-343-361-12	Sequence 12, Appl
100	86	3.1	349	4	US-09-134-001C-4004	Sequence 4004, Ap



Qy	27	APISLGRILIIAGVAGVQVGNALQUSLTLPPYVOTLGLSHALTSPFWMLCGPIAGLVQPL 86
Db	31	AEATLKLGLGVSAVGAQVGNALQUSLTLPPYVQLGLPHTWAAITWLCGPISGMTVQPL 90
Qy	87	VGLYSDRCTSRWRRRRPFILITGMLTICVAVIVVGFSSDITGAALGDTKEHCSLYHGGPRWA 146
Db	91	VGYSDRCTSRFGRRRRPFITGAALVAVAVGLLGFADITGAASGDDPTGWA ---KPR--A 145
Qy	147	AIYVVLGFWLLDFSNTTVQGPARAMADLC--DHHGPSAANSIFCSMMALGNLTGYSGS 204
Db	146	IAVFVVGFWLIDVANNTLQGCPCALLADMAAGSQTTRYANAPSFSPFMAIGNIGGYAAGS 205
Qy	205	TNNHKKWPEFLKTSACEACANLKGAFVLAVVFLVLCLTVTLLFAKE----VPYRANEN 259
Db	206	YSRLTYVPFTKTAACDVYCANLKSCEFTSITLLIVLTIALSVYKEROITDEIQEED 265
Qy	260	LPT--TKAGGEVETPTGPLAVLKGFKDLPMPGSVLLVTAITWLSNFPFLYDTDWMGRE 318
Db	266	LKNRNNSGCGARLPFFGQL--IGALKDLPKMLILLTLTALNIAWFPFLLEFDTDMWGRE 323
Qy	319	IYHCDPKGSNAOISAFNEGVRVCAFGLLNSVTLGFSSFLIEPCKRV--GPRVVVWTSNF 377
Db	324	VY----GGTVGEGKLYDQGVHAGALGLMINSVVLGYMSLSIEGLRMVGAKRLWGIWNI 379
Qy	378	MVCVMAATAALI--SFWSLRDYHGYVDAT--TANASIKAVCLVLVFLAGLVPLAILYSVP 433
Db	380	ILAVCLAMTVLTVTKSAEHFRDSSHIMGSVAPPPPPAGVKGGAIAIEAVLIGPLAITFSIP 439
Qy	434	FANTAOLAAATRGSGOGLCTGVLANISVIPQVIITALCAGPKDALFGKGNIPATGCVASAFAL 493
Db	440	LALATFSASGSGGQSLSLGVLANIAIVVPMFVSITGGMDAMFGGNLPAFVVVGAAT 499
Qy	494	VGVGVGVFLLPKISKQFRAVSAGGH 519
Db	500	ASAVLSFTLLPSPPPRAKIGSGMGH 525

RESULT 4  
US-08-786-555-2  
: Sequence 2, Application US/08786555B  
: Patent No. 5981181  
: GENERAL INFORMATION:  
: APPLICANT: FROMMER, Wolf-Bernd  
: APPLICANT: RIESMEIER, Jörg  
: TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,  
: TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER  
: TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION  
: TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION  
: FILE REFERENCE: Frommer  
: CURRENT APPLICATION NUMBER: US/08/786,555B  
: CURRENT FILING DATE: 1997-01-21  
: EARLIER APPLICATION NUMBER: 08/356,340  
: EARLIER FILING DATE: 1994-12-21  
: EARLIER APPLICATION NUMBER: PCT/EP93/01604  
: EARLIER FILING DATE: 1993-06-22  
: EARLIER APPLICATION NUMBER: DE P4220759.2  
: EARLIER FILING DATE: 1992-06-24  
: NUMBER OF SEQ ID NOS: 11  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 2  
: LENGTH: 525  
: TYPE: prt  
: ORGANISM: Spinacia oleracea  
US-08-786-555-2

Query Match	39.4%	Score 1075;	DB 2;	Length 525;
Best Local Similarity	44.7%	Pred. No. 1.7e-99;		
Matches 226;	Conservative	89;	Mismatches 167;	Indels 24; Gaps 10;
27	APISLGRLLAGMVGAGVQYQWALQSLSTTYPVOTLGLSHALT	SFMMLCGSP	AGLVLVQPL	86
31	AEATLKIKGLVASVAGVQFQWALQSLSTTYPVOLLGPHPTWAAY	ILWLCGSIQMI	VQPL	90

QY 87 VGLYSDRCTSRGRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRWHA 146  
Db 91 VGLYSDRCTSRGRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRWHA 145  
QY 147 ATIVYVGLFWLDFSNNTVQGPARAMADLC--DHGCPAANSIFCSWALGNILYSSGS 204  
Db 146 IAVFVGVFWLDFSNNTVQGPARAMADLC--DHGCPAANSIFCSWALGNILYSSGS 205  
QY 205 TNNHKKWPELTKSACCACANLKGAFVAVVFLVCLTITLIFAKE-----VPRANEN 259  
Db 206 YSLRTVFFFTKTAADVCANLKSCTFISITLLVILALSVVKKRQITIDEIQEED 265  
QY 260 LPT-KTAGEVEETEPTGLPLAVLKGKDPGMPSPVLLVTAITWLSWFFILYDTPMGRE 318  
Db 266 LKNRNNSSCARLPFGQL--IGALKDLPKMLILLVLTALNWIAPFLLDTPDMCKE 323  
QY 319 IYHDPKSNQAISAFNEGVGVAGFLLLNSVLGFSFLETPMCRKV-GRPVVWVTSNF 377  
Db 324 VY-----GCTVGEKGLYDQGVHAGALGLMINSVYLVGMSLSTEGLMVGGAKRLWGI 379  
QY 378 MYCVAMAATALT--SFWSLRDHYGVQDAI--TANASIKAVCLVLFVGLVPLAITYSVP 433  
Db 380 ILAVCLAMTVLTKSAEHRDSSHIMGSAVPPPPAGVGGALAFVGLPIAITFISIP 439  
QY 434 FAVTAQLAATRGGGGLCTGVNLINISIVIPQVITIALGAGPMDALFGKNIPAFGVASAFAL 493  
Db 440 LALASIFASASSGGQGLSLGLVNLALVVPQMEVSVTSGPMDAMEGGNLPAPVVGAVAT 499  
QY 494 VGVVGVVFLLPKISKQRPVAVSAGGH 519  
Db 500 ASAVLSFTLLPSPPPEAKITGSGMGH 525

RESULT 5  
US-09-020-956-113  
; Sequence 113, Application US/09020956  
; Patent No. 6261562  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
; NUMBER OF SEQUENCES: 178  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/020.956  
; FILING DATE: 09-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.427C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 113:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens  
US-09-020-956-113  
Query Match 11.18; Score 303.5; DB 4; Length 553;  
Best Local Similarity 25.38; Pred. No. 6e-22;  
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;  
QY 33 RLILAGVAGVQVQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVVQVPLVGLYSD 92  
Db 17 QLLVNLTLTFCVCLAGITYVPPLLLEVGVEEKFTWVLGIGVGLVCVPLLGSSASD 76  
QY 93 RCTSRWRRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRWHAIVVYL 152  
Db 77 HWRGRRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRWHAIVVYL 127  
QY 153 GFLLDLSNNTVQGPARAMADLC--DHGCPAANSIFCSWALGNILYSSGSTNNWH 209  
Db 128 GVLLDFGVCVCFPLEALLSDFLRDPDH--CROQSVYAFMISLGGCGLYLLPAID--- 182  
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTITLIFAKEVYPRANENLPTTK 264  
Db 183 -WDTSAALPYLGTQEEC-----LFG--LLTLIFT-CVAATLLVAEEAALGPTG--PAEG 231  
QY 265 AGGEVETEPTGLPLAVLKGKDL-----PPGMPSPVLLVTAITWLSWFFILY 310  
Db 232 LSAPSLSPHCCPCARLARLAFRLNGLALLPRLHQLCCRPRTLRLFLVAELCSWMLMTFTLF 291  
QY 311 DTDWNGRIYHGDPK--GSNAQISAFNEGVGVAGFLLLNSVLGFSFLETPMCRKV 367  
Db 292 YTFVGEGLYGVVRAEPGTEAR--RHYDEGVMSGLGFLQCAISLVSFLVMDRLVQRFG 350  
QY 368 PRVWVTSNFMVAVAMAATALISFMSLRDHYGVQDAITANASIKAVCLVLFVGLVPLA 427  
Db 351 TRAVVLSAAFPVAAAGATCL-----SHSVAVVVTASAAITGTFTSALQILPYTLA 400  
QY 428 ILY 430  
Db 401 SLY 403

RESULT 6  
US-09-030-607-113  
; Sequence 113, Application US/09030607  
; Patent No. 6262245  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS  
; NUMBER OF SEQUENCES: 224  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/030.607  
; FILING DATE: 25-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.427C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 113:



SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-09-030-607-113

Query Match 11.1%; Score 303.5; DB 4; Length 553;  
Best Local Similarity 25.3%; Pred. No. 6e-22;  
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;  
OY 33 RLILAGWAGGVOYQWALQSLTTPYVOTLGLSHALTSPFWMCLGPIAGLVQPLVGLYSD 92  
DB 17 QLLVNLITFGLVCLAGITVPPLLLLEVGEVEKFTWVLGIPVLGVPLVGLSASD 76  
OY 93 RCTSMGRRRPFILTCMLICVAVIVGVFSSDIGAALGDTKEHCSLYHGPRWHAIAIVYVL 152  
DB 77 HWRGRRRRPFILWALSGLILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127  
OY 153 GFWLDFSNNTVOGPARMMADLC---DHHGPSAANSIFCSWMLGNILGYSSGSTNNWH 209  
DB 128 GVGLLDFCGQVCFPTPLEALLSDLFDPDH--CROAYSVYAFMISLGGCLGYLLPAID--- 182  
OY 210 KW-----PPELKTSAACCEACANLKGAFVAVVFLVCLTVTLIFAKEVPYRANENLPTTK 264  
DB 193 -WDTSAALPYLGTQEEC-----LFG--LTLFLUT-CVAATLLVAEEAALGPTG--PAG 231  
OY 265 AGGEVETEPTGPLAVLKGFKDL-----PPGMPSVLLVTAITWLSWPPFIY 310  
DB 232 LSAPLSPHCCPCARLARAFNLGALLPRLHQLCCRPRTLRLFLVAELCSWMLMTFTLF 291  
OY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGAAGLLNSVILGFSFLIEPCKRVG 367  
DB 292 YTFDVEGELYGVRAEPGTEAR--RHYDEGVRMGSLGLFLOCAISLVSLVMDRLVQRF 350  
OY 368 PRVWVTNFMVCMVMAATALISFWSLRDYGHVQDATTANASIKAVCLVLFAGLVPLA 427  
DB 351 TRAVLASVAFFPVAAGATCL-----SHSVAVVTASALTGFTFSALQILPYTLA 400  
OY 428 ILY 430  
DB 401 SLV 403

RESULT 7  
US-09-605-785-113  
Sequence 113, Application US/09605785  
Patent No. 6321716  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqi  
APPLICANT: Henderson, Robert A.  
APPLICANT: Kalos, Michael D.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel  
APPLICANT: Wang, Aijun  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Hepler, William  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C16  
CURRENT APPLICATION NUMBER: US/09/605.785

CURRENT FILING DATE: 2000-06-27  
NUMBER OF SEQ ID NOS: 835  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-605-785-113

Query Match 11.1%; Score 303.5; DB 4; Length 553;  
Best Local Similarity 25.3%; Pred. No. 6e-22;  
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;  
OY 33 RLILAGWAGGVOYQWALQSLTTPYVOTLGLSHALTSPFWMCLGPIAGLVQPLVGLYSD 92  
DB 17 QLLVNLITFGLVCLAGITVPPLLLLEVGEVEKFTWVLGIPVLGVPLVGLSASD 76  
OY 93 RCTSMGRRRPFILTCMLICVAVIVGVFSSDIGAALGDTKEHCSLYHGPRWHAIAIVYVL 152  
DB 77 HWRGRRRRPFILWALSGLILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127  
OY 153 GFWLDFSNNTVOGPARMMADLC---DHHGPSAANSIFCSWMLGNILGYSSGSTNNWH 209  
DB 128 GVGLLDFCGQVCFPTPLEALLSDLFDPDH--CROAYSVYAFMISLGGCLGYLLPAID--- 182  
OY 210 KW-----PPELKTSAACCEACANLKGAFVAVVFLVCLTVTLIFAKEVPYRANENLPTTK 264  
DB 193 -WDTSAALPYLGTQEEC-----LFG--LTLFLUT-CVAATLLVAEEAALGPTG--PAG 231  
OY 265 AGGEVETEPTGPLAVLKGFKDL-----PPGMPSVLLVTAITWLSWPPFIY 310  
DB 232 LSAPLSPHCCPCARLARAFNLGALLPRLHQLCCRPRTLRLFLVAELCSWMLMTFTLF 291  
OY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGAAGLLNSVILGFSFLIEPCKRVG 367  
DB 292 YTFDVEGELYGVRAEPGTEAR--RHYDEGVRMGSLGLFLOCAISLVSLVMDRLVQRF 350  
OY 368 PRVWVTNFMVCMVMAATALISFWSLRDYGHVQDATTANASIKAVCLVLFAGLVPLA 427  
DB 351 TRAVLASVAFFPVAAGATCL-----SHSVAVVTASALTGFTFSALQILPYTLA 400  
OY 428 ILY 430  
DB 401 SLV 403

RESULT 8  
US-09-439-313-113  
Sequence 113, Application US/09439313  
Patent No. 6329505  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan Louise  
APPLICANT: Jiang, Yuqi  
APPLICANT: Reed, Steven G.  
APPLICANT: Kalos, Michael  
APPLICANT: Fanger, Gary  
APPLICANT: Retter, Mark  
APPLICANT: Solk, John  
APPLICANT: Day, Craig  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C9  
CURRENT APPLICATION NUMBER: US/09/439.313  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 575  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapien

US-09-439-313-113

```
Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGWAGVQYQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLLTFGLVCLAGITVPPPLLEVGVEEFMTMVLGIGPVGLVCVPLIGSASD 76
QY 93 RCTSRWRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIVVYL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127

QY 153 GFWLDFSNNTVOGPARAMMADLC---DHHGPSAANSIFCSMMALGNILGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTTLIFAKEVPYRANENLPTTK 264
DB 183 -WDTLSALAPYLCTQEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGTE--PAEG 231
QY 265 AGGEVETETGPLAVLKGFKDL-----PPGMPSVLLVTAITWLSWEPFFILY 310
DB 232 LSAPLSLPHCCPCRRARLAFRNLGALLPRLHQLCCRPRTLRLFLVAELCSMMALMTFTLF 291
QY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
DB 292 YTDVFGEGLYQGVPRAEFGTEAR-RHYDEGVRMGSGLFLQCAISLVSFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVVCVMAATALISFWSLRDYGHYVODAITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVYLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILX 430
DB 401 SLV 403

RESULT 9
US-09-352-616A-113
; Sequence 113, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-352-616A-113

Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGWAGVQYQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLLTFGLVCLAGITVPPPLLEVGVEEFMTMVLGIGPVGLVCVPLIGSASD 76
QY 93 RCTSRWRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIVVYL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
```

```
QY 153 GFWLDFSNNTVOGPARAMMADLC---DHHGPSAANSIFCSMMALGNILGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTTLIFAKEVPYRANENLPTTK 264
DB 183 -WDTLSALAPYLCTQEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGTE--PAEG 231
QY 265 AGGEVETETGPLAVLKGFKDL-----PPGMPSVLLVTAITWLSWEPFFILY 310
DB 232 LSAPLSLPHCCPCRRARLAFRNLGALLPRLHQLCCRPRTLRLFLVAELCSMMALMTFTLF 291
QY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
DB 292 YTDVFGEGLYQGVPRAEFGTEAR-RHYDEGVRMGSGLFLQCAISLVSFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVVCVMAATALISFWSLRDYGHYVODAITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVYLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILX 430
DB 401 SLV 403

RESULT 10
US-09-602-877A-101
; Sequence 101, Application US/09602877A
; Patent No. 6432707
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.446C5
; CURRENT APPLICATION NUMBER: US/09/602,877A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-602-877A-101

Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGWAGVQYQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLLTFGLVCLAGITVPPPLLEVGVEEFMTMVLGIGPVGLVCVPLIGSASD 76
QY 93 RCTSRWRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIVVYL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
QY 153 GFWLDFSNNTVOGPARAMMADLC---DHHGPSAANSIFCSMMALGNILGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTTLIFAKEVPYRANENLPTTK 264
DB 183 -WDTLSALAPYLCTQEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGTE--PAEG 231
QY 265 AGGEVETETGPLAVLKGFKDL-----PPGMPSVLLVTAITWLSWEPFFILY 310
DB 232 LSAPLSLPHCCPCRRARLAFRNLGALLPRLHQLCCRPRTLRLFLVAELCSMMALMTFTLF 291
QY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
DB 292 YTDVFGEGLYQGVPRAEFGTEAR-RHYDEGVRMGSGLFLQCAISLVSFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVVCVMAATALISFWSLRDYGHYVODAITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVYLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILX 430
DB 401 SLV 403
```

Db 292 YTFVGEGLYQGVRAEPGTAR-RHYDEGVGMGSLGLFLQCAISLVFSLVMDRLVORFG 350  
 Qy 368 PRVVMVTSNFMVCVAMATAALISFWSLRDHYGYVQDATTANASIKAVCLVLFAFLGVPLA 427  
 Db 351 TRAVYLASVAAPVAAGATCL-----SHSVAVVTASAALTGTFESALQILPYTLA 400  
 Qy 428 ILY 430  
 Db 401 SLY 403  
 RESULT 11  
 US-09-232-149A-113  
 ; Sequence 113, Application US/09232149A  
 ; Patent No. 6465611  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer Lynn  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE  
 ; FILE REFERENCE: 210121.427C6  
 ; CURRENT APPLICATION NUMBER: US/09/232.149A  
 ; CURRENT FILING DATE: 1999-01-15  
 ; NUMBER OF SEQ ID NOS: 338  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 113  
 ; LENGTH: 553  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapien  
 ; US-09-232-149A-113

Query Match 11.18; Score 303.5; DB 4; Length 553;  
 Best Local Similarity 25.38; Pred. No. 6e-22;  
 Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;  
 Qy 33 RLILAGVAGGVQYQWALQLSLTPYVOTLGLSHALTSTFWMVLCGPIAGLVVQVPLVGLYSD 92  
 Db 17 QLLVNLTLFGLVCLAAAGITVPPLLEVGVEEKFTWVIGIPVLGVCVPLLGASD 76  
 Qy 93 RCTSRWGRRRPFIITGCMICVAVIVGVFSSDGAALGDTKEHCSLYHGPRWHAIVVYL 152  
 Db 77 HWRGYGRRRPFIWALSIGILLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127  
 Qy 153 GFWLLDFSNNTVQGPARAMMADLC---DHGPSAANSIFCSMMALGNLTLYSSGSTNNWH 209  
 Db 128 GVGLDFCGQCVFTPLEALLSDLFRDPDH--CROAYSVYAFMISLGGCLGYLLPAID--- 182  
 Qy 210 KW-----FPFLKTSACCEACANLKGAFVAVVFLVLCITVTLIFAKEVPYRANENLPTTK 264  
 Db 183 -WDSALAPYLGTOEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPTE--PAEG 231  
 Qy 265 AGGEVEETPGPLAVLGKDL-----PPGMPSVLLVTAITWLSWFPFITY 310  
 Db 232 LSAPLSLPHCCPCRLARAFNRALLPLHLQCCRMPTLRRLFAELCSMMALMTETLF 291  
 Qy 311 DTDWNGREIYHGDPK---GSAOISAFNEGVVVGAFGLLLNSVILGFSFLLIEPMCKRVG 367  
 Db 292 YTFVGEGLYQGVRAEPGTAR-RHYDEGVGMGSLGLFLQCAISLVFSLVMDRLVORFG 350  
 Qy 368 PRVVMVTSNFMVCVAMATAALISFWSLRDHYGYVQDATTANASIKAVCLVLFAFLGVPLA 427  
 Db 351 TRAVYLASVAAPVAAGATCL-----SHSVAVVTASAALTGTFESALQILPYTLA 400  
 Qy 428 ILY 430  
 Db 401 SLY 403  
 RESULT 12  
 US-09-605-785-708  
 ; Sequence 708, Application US/09605785  
 ; Patent No. 6321716

; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedvick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Hepler, William  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; FILE REFERENCE: 210121.427C16  
 ; CURRENT APPLICATION NUMBER: US/09/605.785  
 ; CURRENT FILING DATE: 2000-06-27  
 ; NUMBER OF SEQ ID NOS: 835  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 708  
 ; LENGTH: 371  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-605-785-708  
 Query Match 8.08; Score 217.5; DB 4; Length 371;  
 Best Local Similarity 26.78; Pred. No. 1.5e-13;  
 Matches 99; Conservative 52; Mismatches 147; Indels 73; Gaps 19;  
 Qy 55 LTPYVOTLGLSHAL---TSFMVLCGPIAGLVVQVPLVGLYSDRCTSRWGRRRPFIITGCM 111  
 Db 10 LVPLPLALYLSQPLTHHTTSLLAGIGPVGLVCPVLGASDHWGRGRRGRRPFIWALSIG 69  
 Qy 112 ICVAVIVGVFSSDGAALGDTKEHCSLYHGPRWHAIVVYVGLFWLLDFSNNTVQGP 171  
 Db 70 ILLSLFLIPRAGWLAGL-----CP---DPRPLELALLILGVLGLDFCGQCVFTPLEAL 120  
 Qy 172 MADLC---DHRGPSAANSIFCSMMALGNLTLYSSGSTNNWHK-----FPFLKTSACCEA 223  
 Db 121 LSDLFRDPDH--CROAYSVYAFMISLGGCLGYLLPAID---WDSALAPYLGTOEC--- 172  
 Qy 224 CANLKGAFVAVVFLVLCITVTLIFAKEVPYRANENLPTTKAGGEVEETPGPLAVLGK 283  
 Db 173 ---LFG--LLTLIFLT-CVAATLLVAEEAALGPTE--PAEGLSAPLSLPHCCPCRLAR 224  
 Qy 284 KDL-----PPGMPSVLLVTAITWLSWFPFITYDWMGREIYHGDPK---G 326  
 Db 225 RNLGALLPRLHLQCCRMPTLRRLFAELCSMMALMTETLFYDVFVGEGLYQGVRAEPG 284  
 Qy 327 SNAOISAFNEGVVVG-----FGLLLNSVILGFSFLLIEPM-----CRKVGPRV 371  
 Db 285 TEAR-RHYDEKALAAASRGWCGSRPPETTLGAVSLV-PLHPGPDFSVKVGMDPIC 342  
 Qy 372 --WYTSNPMVC 380  
 Db 343 PSWVW-NISAC 352  
 RESULT 13  
 US-08-595-553A-2  
 ; Sequence 2, Application US/08595553A  
 ; Patent No. 6391579  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NANCY CARRASCO, ET AL.  
 ; TITLE OF INVENTION: THYROID SODIUM/IODIDE SYMPORTER AND  
 ; FILE REFERENCE: NUCLEIC ACID ENCODING SAME  
 ; NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
 STREET: 90 PARK AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: U.S.A.  
 ZIP: 10016  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE  
 MEDIUM TYPE: DISKETTE  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/595,553A  
 FILING DATE: FEBRUARY 1, 1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CRAIG J. ARNOLD  
 REGISTRATION NUMBER: 34,287  
 REFERENCE/DOCKET NUMBER: 96700/393  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 697-5995  
 TELEFAX: (212) 286-0854 or 286-0082  
 TELEX: TWX 710-581-4766  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 618  
 TYPE: AMINO ACID  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE:  
 DESCRIPTION: PROTEIN  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: RAT  
 INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER  
 US-08-595-553A-2

Query Match 4.6%; Score 126.5; DB 4; Length 618;  
 Best Local Similarity 21.7%; Pred. No. 0.00043;  
 Matches 121; Conservative 71; Mismatches 210; Indels 155; Gaps 25;

QY 1 MARG-----DGGQLAELSGVGAAGAAVVDHVPISLGRLLIAGWAGVGYG--- 47  
 DB 36 LARGGORSADFFTGGRQLAAAPVGVLSLAASFMSAVQ-----VLGVPAEARYGLKF 87  
 QY 48 -W-ALQL-SLLT-----PYVOTLGLSHA-----LTSPMMLCGTAGLVVQPLVGLY 90  
 DB 88 LHWACQQLNLSLLTAFLELPFYRGLTSTYOYLELRFSAVRLCTQYLV----- 140  
 QY 91 SDRCTSRWGRRRPFIITGCMILICAVI---VVGFSDDIGAALGDTKEHCSLYHGPRWHA 147  
 DB 141 -----TMLYTGIVIVAPALILNQVTGL-DIWASLSTGIITCTLTVTYVGMKA 186  
 QY 148 IYV-----VLGFLLDFSNNTVQPARAMMA-----DLCDDHGPSAANSIFCS 190  
 DB 187 VVMTDVFQVVMVLGVFVILARGVILGGPRNVLSLAQNHRSINLMDFFDPDRSRYITFT 246  
 QY 191 WMAIGNILYSGSGSTNNHKKFPFLKTSACCEACANLKGAFV--AVFVLVCLVTLIF 248  
 DB 247 FIVGGTLVWLSMYGVNOAQ-----VORYVACHTEGAKALLVNLQGLFLIVASAACCGI 301  
 QY 249 AKEVPYRANENLPTTKAGGEVETETGPLAVLKGFDLPDPPGMPSVLLVTATWLSWFFPI 308  
 DB 302 VMFVYKDCDPLTGRISA---PDQYMLPLVLDIFEDL-PGVPGFLACA----- 347  
 QY 309 LYDTHMGREIYHGDPKGSNAQISAFNEGVVRGAFGLLLNSVILGFSSFLPEPMCRKVG 368  
 DB 348 -----YSGTISTASTSINAM-AAVT-----EDLIKPRMPGLAP 380  
 QY 369 R-VVWVTSNF-----MVCVMAATAALI-----SEWLSRDYHGVQDAITANASIKAV 414  
 DB 381 RKLIVISKLSFIYGSACLTVAALSSLLGGVGLQSGFTVMGVISGPLLCAFTLGLMLPA- 439

QY 415 CLVLFALFGLVPLAILYSVPFAVTAQL-----AATRGGGGLCTGVNLINISIVPQVIALGA 470  
 DB 440 CNTFCVLSGLAAGLAVSLWAVGATLYPPGEGTMCGLVLTSAAGCTNDSVLL----- 490  
 QY 471 GPWDALFGKGNIPAFGV 487  
 DB 491 GPPGATNASNGIPSSGM 507  
 RESULT 14  
 5268463-9  
 ; Patent No. 5268463  
 ; APPLICANT: JEFFERSON, RICHARD A.  
 ; TITLE OF INVENTION: PLAANT PROMOTER a-GLUCURONIDASE GENE  
 ; CONSTRUCT  
 ; NUMBER OF SEQUENCES: 9  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/447,976  
 ; FILING DATE: 08-DEC-1989  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 119,102  
 ; FILING DATE: 10-NOV-1987  
 ; APPLICATION NUMBER: 264,586  
 ; FILING DATE: 31-OCT-1988  
 ; SEQ ID NO:9  
 ; LENGTH: 435  
 5268463-9

Query Match 4.2%; Score 114; DB 6; Length 435;  
 Best Local Similarity 19.0%; Pred. No. 0.0046;  
 Matches 92; Conservative 61; Mismatches 161; Indels 170; Gaps 18;

QY 43 GVOYCWALQLSLLTPYVOTLGLSHALTSMFLCGLPIAGLVVQPLVGLYSDRCTSRWGR 102  
 DB 17 GIVY-----MYLMYYTDDVGLSVGLVGLTFLVARIMDAINDPIMPINGVIVNATRSRWGR 71  
 QY 103 PFILTCGM-----LICVAVIVVG-----FSSDIGAALGD 131  
 DB 72 PWILIGTLANSVILLFLFSAHLEFEGTQIVFVCTVYILWGMTYTIMDIPFWSLVPTILD 131  
 QY 132 TREHCSLYHGPRWHAIV-YVLGFLLDFSNNTVQPARAMMADLCDHGPSAANSIFCS 190  
 DB 132 KRERQLVPYPRFFASLAGFVTAGVTLFPVN----- 162  
 QY 191 WMAIGNILYSGSGSTNNHKKFPFLKTSACCEACANLKGAFVAVVFLVCLVTLIFAK 250  
 DB 163 -----YVGGG---DRGFGQMT-----LVLIATFIVTITILRNHV 196  
 QY 251 EYVYRANENLPTTKAGGEVETETGPLAVLKGFDL---PPGMPSVLLVTATW----- 301  
 DB 197 EV-FSSDNQPSAE-----GSHLTAKAIVAIYKNDQLSCLGLMALAYNVASNI 244  
 QY 302 ---LSWFPFETLYDTHMGREIYHGDPKGSNAQISAFNEGVVRGAFGLLLNSVILGFSS 357  
 DB 245 TGFALYYSYVIGDADLPYLYSY-----AGANLV---TILVFFPR 283  
 QY 358 LIETPCMRKVGPRVVMVTNFMVCMVMAATAALISFWSLRDHYGVQDAITANASIKAVCLV 417  
 DB 284 LVKSLSR---RILWAGASILPVLSCGVLLLM---ALMSYHNV--LIVAGILLNVGTA 334  
 QY 418 LFAFLGVPL-----ATLYSVFPFAVTAQLAATRGGGGLCTGVNLINISIV 460  
 DB 335 LFWLQVLMVADIVDGYEYKLVRCESLAYSV-----QTMVVKGSFAFFIAVVLGMICY 388  
 QY 461 IPOV 464  
 DB 389 VPNV 392  
 RESULT 15  
 US-09-134-001C-3020  
 ; Sequence 3020, Application US/09134001C  
 ; Patent No. 6380370

```

: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3020
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3020

```

```

Query Match      4.18; Score 112; DB 4; Length 467;
Best Local Similarity 22.0%; Pred. No. 0.0081;
Matches 115; Conservative 56; Mismatches 179; Indels 172; Gaps 25;

```

```

Qy 28 PISGLRLTLAGWAGGV-QYGMALQLSLTPYVOTLGLSHALTSFWM-LCGPIAGLVVQP 85
Db 57 PISL-----INGFVALVMTGGIELAFSLRYVVDLGFSPQATLLFTYVGLFAI----- 108
Qy 86 LVGLYSDRCTSRWGR---RRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHCS-LYHG 141
Db 109 -----SSWGSGLVAETFGPRIMLIGVAWIVFVFLFLFGLFLEQOXYTLWVYFG 158
Qy 142 PRWHAIVVYVIGF--WLDFSNNTVQGPARAMADLCHDHGSPSAANSIFCSWMAIG-NIL 198
Db 159 IRGLAYPLFIYAFMWIAQVT-----PGARM-----ASAMGWFSWYMCIGILL 202
Qy 199 GYSSGSTNNHKKWPFLLKTSACCACANLKGAFLVAVVFLVLC-----TVTL 246
Db 203 G-----NWIPSLISIRGIFNTLWGVFWFVAVAGLMIMYLKRGAGKPDPAVTL 252
Qy 247 IFAKEVPYRANENLPTTKAGEVETEPTGPLAVLKGFKDLPPGMPSVLLVTAITWLSWF- 305
Db 253 V-----ERLKLSSGVITIAER-----GMFLVLVRLICNLSLFG 288
Qy 306 -PFILYDPMWGREIYHGDPKGNAQISAFNEGVRVGAFLGDLNSVLGFS----- 355
Db 289 LPVIL-----PLYLTSVEVGFSGMEQWLHWG 314
Qy 356 -----SFLIEPMCRKVGPRVW-VTSNFMVCMVAMAATALISFWSLSDYHGIVQDAITANA 409
Db 315 VMFVSIFTNVIWGOIGDRGLWLIOMRFGCIGCALSSL-APFYLPQFYG-----AHF 366
Qy 410 SIKAVCLVFLAFGLVPLAIXSVPF-AVTAQLAATRG-----GOCCLCTGVNLNISIVIPQV 464
Db 367 EI-----ALIAAIGGISVTAFPVPGAVFLALAPEQGAALSAHNLAAGLSNF----- 414
Qy 465 IIALGAGPMDALFKGNTPAFGVSASFA--LVGGVGVVFLLP 504
Db 415 -----MGPGIATLTIATLIGKGVVMIYAGLVVAGAVLTFPI 451

```

```

RESULT 16
US-08-403-852D-21
: Sequence 21, Application US/08403852D
: Patent No. 5891695
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Blanche, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: APPLICANT: De Crecy-Lagard, Valerie

```

```

: TITLE OF INVENTION: Polypeptides Involved In The
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/08/403,852D
: APPLICATION NUMBER:
: FILING DATE: 10-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR 93/00923
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/11441
: FILING DATE: 25-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Meyers, Kenneth J.
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 03806.0054-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 528 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-403-852D-21

```

```

Query Match      4.18; Score 112; DB 2; Length 528;
Best Local Similarity 20.1%; Pred. No. 0.0097;
Matches 103; Conservative 62; Mismatches 149; Indels 198; Gaps 27;

Qy 91 SDRCTSRGRRRPFILTCMLICVAVIV-----GFSSDIGAALGDTKEHCSL 138
Db 28 ADPRRW-----VILG--VICLAQLVLDNTVNLVAVIPVLTDLGASTADIO----- 74
Qy 139 YHGPWHAAIYVYVIGFWLLDFSNNTVQGPARAMADLCHDHGSPSAANSIFCSWMAIGNL 198
Db 75 -----NMN-AYALVQSGLLTAGSLADRYGRKRL-----LMLGLVL 110
Qy 199 GYSSGSTNNHKKWPFLLKTS-----CCEACANLKGAFLVAVVFLVLCVTLIFAKEVPYR 255
Db 111 -FGAGSA-----WAAFAQDSAQLIAARAGMGVGGALLATTTLAV-----IMQVFDDEPR 160
Qy 256 ANENLPTTKAGEVETEPTGPL-----AVLKG-----KDLPPGMPSVLLVTAITWLSWF 305
Db 161 AIGLWGAASSLGFA-----AGPLLGALLDHFWSGIFLINLPVALLGLLAVLV----- 211
Qy 306 PFILYDPMWGREIYHGDPKGNAQISAFNEGVRVGAFLGDLNSVL-ILGFSFLIEPMCR 364
Db 212 -----PETKNP-----EGRRPDLGLAVLSTLGMGVVVAIIS----- 243
Qy 365 KVGRVWVWTSNFMVCMVAMAATALISF--WSLRDYH----- 398
Db 244 --GPEHGWTAPQVLLPAAVAAALATFVRWELHTPHMLDMGFFTDRENGSPAECSF 301
Qy 399 -----CYVODAI-----TANASKAVCL-----VLFAFLGVPLAIL 429
Db 302 GMAGSLFLLTQHLQVLGY--DALQAGLRTAPLATIIVALNLAGLAKLALGATARSIA 359
Qy 430 YSVPPAVTAQLAATRGCGQ-----GLCTGVL-----NISIVIPQVIAL----- 468

```

Db 360 LGMTL-LAAGLSAVAVGGSPDAGYGGMLAGLLMGAGIALAMPAMATAVMSSIPPAKAG 418  
QY 469 -GAGPMDALFGKGNIPAFGVASAFALVGGVVG 499  
Db 419 AGAGV-----OCTLTFEGGGLGVALGVLG 444

RESULT 17  
US-08-510-646B-22  
:Sequence 22, Application US/08510646B  
: Patent No. 607699  
: GENERAL INFORMATION:  
: APPLICANT: Blanc, Veronique  
: APPLICANT: Blanc, Francis  
: APPLICANT: Crouzet, Joel  
: APPLICANT: Jacques, Nathalie  
: APPLICANT: Lacroix, Patricia  
: APPLICANT: Thibaut, Denis  
: APPLICANT: Zagorec, Monique  
: APPLICANT: Debussche, Laurent  
: APPLICANT: De Crecy-Lagard, Valerie  
: TITLE OF INVENTION: Polypeptides Involved In The  
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
: TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
: NUMBER OF SEQUENCES: 45  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
: STREET: 1300 I Street, N.W., Suite 700  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: USA  
: ZIP: 20005-3315  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/510,646B  
: FILING DATE: 03-AUG-1995  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/403,852  
: FILING DATE: 10-MAY-1995  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/FR 93/00923  
: FILING DATE: 25-SEP-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: FR 92/11441  
: FILING DATE: 25-SEP-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Meyers, Kenneth J.  
: REGISTRATION NUMBER: 25,146  
: REFERENCE/DOCKET NUMBER: 03806.0054-01000  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202) 408-4000  
: TELEFAX: (202) 408-4400  
: INFORMATION FOR SEQ ID NO: 22:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 528 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
US-08-510-646B-22

Query Match 4.1%; Score 112; DB 3; Length 528;  
Best Local Similarity 20.1%; Pred. NO. 0.0097;  
Matches 103; Conservative 62; Mismatches 149; Indels 198; Gaps 27;

QY 91 SDRCTSRWRRRPFILTCMLICVAVIV-----GFSSDIGAALGDTKEHCSL 138  
Db 28 ADRPRW-----VILG--VICLAQLVLLDNTLVNVAIPVLTDLGASTADIQ----- 74

QY 139 YHGRWHAIVVVLGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSIFCSWMALGNIL 198  
Db 75 -----WMIN-AYALVQSGLLLTAGSLADRYGRKRL-----LMLGLVL 110  
QY 199 GYSGSTNNWHKWFPLKTS-----CCEACANLKGAFLVAVVFLVCLTVTLIFAKEVYR 255  
Db 111 -FGAGS-----WAAFAQDSQAIIAARAGMGVGGALLATTITLAV-----IMQVDDDERPR 160  
QY 256 ANENLPTTKAGEVETPTGPL--AVLKG-----KDLPPGMPSVLLTYTITWLSWF 305  
Db 161 AIGLWGAASSLGFA-----AGPLLGALLDHFWMGSIIFINLPVALLGLLAVARLV----- 211  
QY 306 PFILYDTDMGRIYHGDPKGNSQAISAFNEGVRVGAFGLLINSV-ILGFSFSLIEPMCR 364  
Db 212 -----PETKNP-----EGRRPDLGLGAVLSTLGMVGVYVAILIS----- 243  
QY 365 KVGRVWVWTSNFMVVCVMAATALISF--WSLRDYH-----VLFALGVPLAIL 429  
Db 244 -GPEHGTAPQVLLPAAVAAALTAFAVRWELHTPHMLDMGFFTDRRFNGSPSPAECSF 301  
QY 399 -----GYVODAI-----TANASIKAYCL-----VLFALGVPLAIL 429  
Db 302 CMAGSLFLLTQHLQVLVGY--DALQAGLRTAPLALTIVALLNLAGLAKLLAALGTARSTIA 359  
QY 430 YSVPEAVTAQAATRGQ-----GLCTGVL-----NISVIVPOVILAL----- 468  
Db 360 LGMTL-LAAGLSAVAVGGSPDAGYGGMLAGLLMGAGIALAMPAMATAVMSSIPPAKAG 418  
QY 469 -GAGPMDALFGKGNIPAFGVASAFALVGGVVG 499  
Db 419 AGAGV-----OCTLTFEGGGLGVALGVLG 444

RESULT 18  
US-09-231-818-21  
: Sequence 21, Application US/09231818  
: Patent No. 6171846  
: GENERAL INFORMATION:  
: APPLICANT: Blanc, Veronique  
: APPLICANT: Crouzet, Joel  
: APPLICANT: Jacques, Nathalie  
: APPLICANT: Lacroix, Patricia  
: APPLICANT: Thibaut, Denis  
: APPLICANT: Zagorec, Monique  
: APPLICANT: Debussche, Laurent  
: APPLICANT: De Crecy-Lagard, Valerie  
: TITLE OF INVENTION: Polypeptides Involved In The  
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
: TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
: NUMBER OF SEQUENCES: 43  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
: STREET: 1300 I Street, N.W., Suite 700  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: USA  
: ZIP: 20005-3315  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/231,818  
: FILING DATE: 03-AUG-1995  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: FR 92/11441  
: FILING DATE: 25-SEP-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Meyers, Kenneth J.  
: REGISTRATION NUMBER: 25,146  
: REFERENCE/DOCKET NUMBER: 03806.0054-01000  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202) 408-4000  
: TELEFAX: (202) 408-4400  
: INFORMATION FOR SEQ ID NO: 22:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 528 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
US-09-231-818-21

APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-00000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 528 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-231-818-21

Query Match 4.18; Score 112; DB 4; Length 528;  
 Best Local Similarity 20.18; Pred. No. 0.0097;  
 Matches 103; Conservative 62; Mismatches 149; Indels 198; Gaps 27;  
 QY 91 SDRCTSRGRRRRPILGCMILICVAVIV-----GFSSDICAALGDTKEHCSL 138  
 DB 28 ADPRRW-----VILG--VICLAQLVLDNTVLNVAIPVLTTDLGASTADIQ----- 74  
 QY 139 YHGRPHAAIYVVLGFWLLDFSNNTVQPARAMMADLCHDHGPSAANSIFCSMMALGNIL 198  
 DB 75 -----WMIN-AYALVQSGLLTAGSLADRYGRKRL-----LMLGLVL 110  
 QY 199 GYSGSTNNHKKWPFPLKTS-----CCACANLKGAFVAVVFLVCLTVLIFAKEVPYR 255  
 DB 111 -FGAGSA-----WAAFAQDSQAIAARAGMGVGGALLATTTTAV-----IMQVDDDERPR 160  
 QY 256 ANENLPTTKAGEVETETGPL-----AVLKGK-----KDLPPGMPVSVLLVTAITWLSWF 305  
 DB 161 AIGLWGAASSLGFA-----AGPLLGALLDHFWSGIFLINLPVALLGLLAVARLV----- 211  
 QY 306 PFILYTDWGMREIYHGDPKGNSAQISAFNEGVRVGFAGLLNSV-ILGFSSFLIEPMCR 364  
 DB 212 -----PETKNP-----EGRRPDLGLGAVSLTGMVGVVYAIIS----- 243  
 QY 365 KVGPRVWVTSNFMVCMAMATALISF--WSLRDYH----- 398  
 DB 244 --GPEHGTAPQVLLPAVAALTAFAVRWELHTPHPMLOMGFFTDRENFSPSPAECSSE 301  
 QY 399 -----GYODAI-----TANASIKAVCL-----VLFAFLGVPLAIL 429  
 DB 302 GMAGSLFLLTQHLQVLGY--DALQGLRTAPLALTIVALLMAGLAKLAAAGTAKSIA 359  
 QY 430 YSVFPAVTAQLAATRGGGQ-----GLCTGVL-----NISIVIPQVITAI 468  
 DB 360 LGMTL-LAAGLSAVAVGSGPDAGYGGMLAGLLMGAGIALAMPAMATAVMSSIPPAKAG 418  
 QY 469 -GAGPDALFGKGNIPAFGVASAFALVGGVVG 499  
 DB 419 AGAGV-----QGTLEFGGGLGVAILGAVILG 444

RESULT 19  
 US-08-882-704A-6  
 Sequence 6, Application US/08882704A  
 Patent No. 5879906  
 GENERAL INFORMATION:  
 APPLICANT: Jefferson, Richard A.  
 APPLICANT: Wilson, Katherine J.  
 APPLICANT: Leader, Michael  
 TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED AND BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington

COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/882,704A  
 FILING DATE: 25-JUN-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5879906tenburg Ph.D., Carol  
 REGISTRATION NUMBER: 39,317  
 REFERENCE/DOCKET NUMBER: 190106.404  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 457 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-882-704A-6

Query Match 3.98; Score 106; DB 2; Length 457;  
 Best Local Similarity 20.88; Pred. No. 0.032;  
 Matches 103; Conservative 58; Mismatches 205; Indels 130; Gaps 18;  
 QY 36 LAGWAGVQYQWAL----QLSLLTPYVQTGLSLHALTSFMMLCGPIAGLVVQPLVGLYS 92  
 DB 10 IVGYSLGDVANNFAFAGALFLLSVYTDVAGVGAAGTALLVRYVEDAFADVAGRVVD 69  
 QY 93 RCTSRGRRRRPILTG-CMLICVAVIVVGFSSDICAALGDTKEHCSLHGRPHAAIYV 151  
 DB 70 SVNTRGKFRPFLFTGAPLMIFSVLVFWPTDWS-----HGSKVYVAYLY 116  
 QY 152 LGFWLLDFSNNTVQPARAMH-----ADLCHHGPSAANSIFCSMMALGNILGYSSGS 204  
 DB 117 MGLGLCYSLNIPYGSLATANTQOQSPARLGAARGIAASLTFVCLAFGLGSPKNSPE 176  
 QY 205 --TNNHKKWPFPLKTSACCEACANLKGAFVAVVFLVCLTVLIFAKEVPYRANENLPT 262  
 DB 177 EMVSVYHFWITVLA-----IAGMVLFCFKSTRENVRIVAP---SLNISLOT 223  
 QY 263 TKAGEVETETGPLAVLKGFKDLPPGMPVSVLLVTAITWLSWFFILYDTHWGMREIYHG 322  
 DB 224 LKRRPLFMLCIGALCVL-----ISTFAVSASS-LFYVRYVLDNLTGLTFLVLVQ 272  
 QY 323 DPKGNSAQISAFNEGVRVGAFLGLLNSVILGFSSFLIEPMCRKVGPRVWVTSNFMV 381  
 DB 273 NLVGTVASAPLVPMVARIGKKNFTLIGALLGTCGYLL-----FFWV-SVWSLPV 321  
 QY 382 AMAATALIS-----FWSLR---DYHGYVQDAITANASIKAVCLVLFALFGLVPLAI 428  
 DB 322 ALVALATAISGQGVMTVMWALEADTVEYGEYL-----TGVRIGLTYSLPSF----- 369  
 QY 429 LYSVFFAVTAQLAATRGGGQGL-----CTGVNLNISIVIPQVITAIAGAPWDALFG 478  
 DB 370 -----TRKCGAIGGSIAPFILGLSGYIANQVOTPEVIN----- 403  
 QY 479 KGNIPAFGVASAFALV 494  
 DB 404 -----GIRTSIALV 412

RESULT 20  
 US-09-151-957-6  
 Sequence 6, Application US/09151957  
 Patent No. 6429292  
 GENERAL INFORMATION:  
 APPLICANT: Jefferson, Richard A.





```

; GENERAL INFORMATION:
; APPLICANT: Aaron Kaplan et al.
; TITLE OF INVENTION: ENHANCING INORGANIC CARBON FIXATION BY
; TITLE OF INVENTION: PHOTOSYNTHETIC ORGANISMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2.
; SOFTWARE: Word for Windows version 3.11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332.041
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 325/45
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-332-041-5

```

```

Query Match 3.8%; Score 103.5; DB 4; Length 474;
Best Local Similarity 22.1%; Pred. No. 0.059;
Matches 60; Conservative 45; Mismatches 85; Indels 81; Gaps 17;

Qy 26 VAPISLGLILAGWVA-----GGVQY---GWA---LQLSLTPPVQTLGLSHALTSEW 73
Db 2 VSPISIRSLMFGFSPQEWGRGSLVRLVGVGQSWIQASVLWPHFALGT--ALVAIIF 59

Qy 74 LCGPIAGLVQPLVGLYSDRCTSRGRRRPFILTGC-----MLI-----CVAVIV 118
Db 60 IAAPTSTT---MLGIFMLCCAFWA-----LTFADQPGKGLTPIHVLVFAVWICISAIA 111

Qy 119 VGFSS-DTGAALGDTKEHCS-----LYHGPRW---HAAVYVGLFWLDFE-NNTV 164
Db 112 VGFSPVMAAASGLAKLTANLCLFLAARLLQNKQWLNRLVTVLLVGLLVGSYGLRQOV 171

Qy 165 QPARAMADLCHHGCPA-----ANSIFCSWALGNIL-----GYSSGSTNHHWKF 212
Db 172 DG-----VEQLATWNDPTSTLAQATRVY-SFLGNPNLLAAYLVPMPTGLSLSALVYVRRW 225

Qy 213 PFLKTSACCEACANLKGAFLVAVVFLVCLT 243
Db 226 P-----KLLGATWIVNLLCLFFT 244

```

```

RESULT 23
US-09-031-392-2
; Sequence 2, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.

```

```

; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031.392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 563 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-031-392-2

```

```

Query Match 3.7%; Score 101; DB 2; Length 563;
Best Local Similarity 19.1%; Pred. No. 0.14;
Matches 113; Conservative 78; Mismatches 214; Indels 186; Gaps 29;

Qy 4 GDGGGLAE---LSAGVGAAGVAVDHVAPISLGRLLIAGWAGGVQYGNALQ-LSLLTPV 59
Db 28 GPGRALLECDHLRSGVGGRRRKKDWSCSLVAS--LAGAFSGSFLYGNLVNVPATPYI 85

Qy 60 Q-----TGLSHALTSEFMVLCPIAGLVQPLVGLVSDRCTSRWGR 101
Db 86 KAFYNESWERRHGRPIDPTLTLLWSVTVSIFAIGLVGLTLVKNIG-----KVLGRK 138

Qy 102 RPFILTCMLICVAVVGVGFSDDIGAALGDTKEHCSLYHP-----RWHAAI-----V 149
Db 139 HTLLANNGFATSAALLMA-----CSLQAGAFEMLVGRFIMGDGVVAL 182

Qy 150 YVLGFWLDFSNNTVQGPARAMMADLCHDHGSPSAANSIF-CSWMAIGNILGYSS--GSTN 206
Db 183 SVLPYMLSEISPKIRGSLGVQVTA-----IFICIGVFTGOLLGPELLGKES 229

Qy 207 NWHKWFPEFLKTSACCEACANLKGAFLVAVVFLVCLT---LIFAKEVPYRANENL 260
Db 230 TW-----PY-----LFGVIVVPAVQVLLSLPFLPSRVLLEKHNARAVKAF 273

Qy 261 PTTKAGGVEVETPTGPLA---VLKGFKDLPPGMPSVL-----LVTAITWLS--- 303
Db 274 QTFGLKADVSQVEEVLAEHSVORSIR-----LVSVELELLRAPYRVQWVTVIVTMACVQ 328

Qy 304 -----WFPFLYDTHWGREIYHGDPKGSAQNAISAFNEGVVGVCAFGLLLNSVILGRSS 356
Db 329 LCGLNAIWF-----YTNISIFGKA--GIPPAKIPYVTLSTGGI-----ETLAAVFSG 372

Qy 357 FLIEPWCRCVGRVVRVVTSTNFMVCMVMAATALISFWSLRDHYGVQDATTANASIKAVCL 416
Db 373 LVIEHLGR--PLLI---GGFGLMGLFTGLTITL-TLQDHAPV-----PYLSIVGILA 421

Qy 417 VLFAGLVPLAIVSVPEAVTAQL--AATRGGGGLCTGVNLISIVIPQVITIALGAGPWD 474

```

```

Db 422 IIAFSCGP-----GGIPFILTGEFFQSOR-----PAAFIAGTVNWL 460
QY 475 ALFGKG-----NIPAFGVASAFALVGGVGVGVELLPKISKRFRAVS 515
Db 461 SNFAVGLLFPFIQKSLDTCYCLVFATCITGAIYLYFVLPETKNRTYAEIS 511

RESULT 24
US-09-299-549-2
; Sequence 2, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549
; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 563 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-299-549-2

Query Match 3.7%; Score 101; DB 4; Length 563;
Best Local Similarity 19.1%; Pred. No. 0.14;
Matches 113; Conservative 78; Mismatches 214; Indels 186; Gaps 29;

QY 4 GDGGQLAE---LSAGVGRGAADVHVAPISLGRLLIAGVAGVOYGNALQ-LSLLTPVY 59
Db 28 GGRALLECDHLRSGVPGGRKKDWSCLVAS--LAGAFSGSFYLYGNLSVNNAPTPIY 85
QY 60 Q-----TLGLSHALTSFMWLCGPIAGLVVQPLVGLVSDRCTSRWGR 101
Db 86 KAFYNESWERRHGRPIDPTLLWSVTVSIFAIGLVGLTVKMG-----KVLGRK 138
QY 102 RPFILTCMLICVAVIVGVFSSDIGAALGDTKEHSLYHGP-----RWHAAI-----V 149
Db 139 HTLLANNGFSAALIMA-----CSLQAGAFEMLIYVGRFIMGIDGGVAL 182
QY 150 YVLGFLLDFSNNTVCGPARAMADLCDHHGFSAAANSIF-CSWMAIGNILGYSS--GSTN 206
Db 183 SVLPWYLSISPKRINGSLGQVTA-----IPICIGVTGQLGLGPELLGNES 229

```

```

QY 207 NWHKWFPELKTSAACEACANLKGAFVAVVFLVLCITVT-----LIFAKEVYPYRANENL 260
Db 230 TW---PY-----LFGVIVVPAVQQLSLPFLPDSRYLLEKHNEARAVKAF 273
QY 261 PTTKAGGEVETPTGFLA---VLKGFKDLPGCMPSVL-----LVTAITWLS--- 303
Db 274 QTFLGKADVSGVEEVLAESHVQSRIR-----LVSVELLRAPYVRWQVTVVITMACYQ 328
QY 304 -----WPFILYDIDWGREIYHGDPKGSNAQISAFNEGVVRGVGAFGLLLNSVILGFSS 356
Db 329 LCLGNAINF-----YTNSIFGKA--GIPPAKIPYVTLSTGGI-----ETLAASFSG 372
QY 357 FLIEPMCKRVGPRVYVVTNFMVVCVMAATALISFWSLRDHYGYQDAITANASIKAVCL 416
Db 373 LVIEHLGRR--PLLI---GGFGLMGLFTGTITL-TLQDHAPVW-----PVLISVIGILA 421
QY 417 VLFAPLGVPLAILYSVPAVTAQL--AATRGGGGLCTGVNLNISIVIPQVIALGAGPWD 474
Db 422 IIAFSCGP-----GGIPFILTGEFFQSOR-----PAAFIAGTVNWL 460
QY 475 ALFGKG-----NIPAFGVASAFALVGGVGVGVELLPKISKRFRAVS 515
Db 461 SNFAVGLLFPFIQKSLDTCYCLVFATCITGAIYLYFVLPETKNRTYAEIS 511

RESULT 25
US-09-610-417-2
; Sequence 2, Application US/09610417
; Patent No. 6346374
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/610,417
; FILING DATE: 05-Jul-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/299,549
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 563 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
; US-09-610-417-2

Query Match 3.7%; Score 101; DB 4; Length 563;
Best Local Similarity 19.1%; Pred. No. 0.14;

```

```
Matches 113; Conservative 78; Mismatches 214; Indels 186; Gaps 29;
Qy 4 GDGGQAE--LSAGVRCAAAVDVHAPISLGRLLLAGWAGGVQYXWALQ-LSLLTPYV 59
Db 28 GPRALLECDHLRSGVPGRRRKQWSSCLLVAS--LAGAFSSFLYGYNSVWNPAPYI 85
Qy 60 Q-----TLGSLHALTSFWMLCGPIAGLVVQPLVGLYSDRCTSRMGR 101
Db 86 KAFYNESWERRHGRPIDPDTLTLLWSVTVSIFAIGGLVGTLLVKMIG-----KVLGRK 138
Qy 102 RPFILTCMLICVAVIVGVSSDIGALGDKHCSLYHGP-----RWHAAI-----V 149
Db 139 HTLLANNGFALSAALLMA-----CSLQAGAFEMLIIVGRFIMGIDGGVAL 182
Qy 150 YVLGFWLLDFSNNTVQGPARAMMADLCDHGHGSAANSIF-CSWMALGNILGYSS--GSTN 206
Db 183 SVLPWYLSSEISPKIRSLGQVTA-----IFICIGVFTGOLLGLPELLGKES 229
Qy 207 NWHKWFPELKTSAACEACANLKGAFVAVFVLCLVT-----LIFAKEVPYRANENL 260
Db 230 TW----PY-----LFGVIVPFAVVQLLSLPPLDPSPRYLLEKHNEARAVKAF 273
Qy 261 PTTKAGGEVETPTGCLA---VLKGFCDLPGMPSVL-----LVTAITWLS--- 303
Db 274 QTFLGKADVSOEVEVLAEASHVQRSIR-----LVSLELLRAPYVRWQVTVIVTMACYQ 328
Qy 304 -----WFPEILYDTMMGREIVHGDPKGSNAOISAFNEGVRVGAFGLLLNSVILGFSS 356
Db 329 LCGLNAIWF-----YTNSTFGKA---GIPPAKIPVTLSTGGI-----ETLAAVPSG 372
Qy 357 FLIEPMCKRVGPRVWVTSNFMVCMVAMAATALISFWSLRDYHGYYQDAITANASIKAVCL 416
Db 373 LVIEHLGRR--PLLI---GGFGLMGLRFGTLITL-TLQDHAPWV-----PYLSIVGILA 421
Qy 417 VLFAPLGVPLAILYVFPVAVTAQL--AATRCGGGGLCTGVLNLSIVIPQVILALGAGPWD 474
Db 422 IIAFCSCGP---GGIPFILTGEFFQOSQR-----PAAFIIAGTVNWL 460
Qy 475 ALFGKG-----NIPAFGVASAPALGVGVGVFLLPKISKROPRAYS 515
Db 461 SNFAVGLLPFIQKSLDTCFLVATICITGAIYLVFVLPETKNTYAEIS 511
```

Search completed: April 17, 2003, 11:33:53  
Job time : 23 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:30:45 ; Search time 19 Seconds  
(without alignments)  
2065.453 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSAVGRAA.....VFLPKISKRQRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 288829 seqs, 7561385 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	303.5	11.1	553	9	US-09-232-880-113
2	303.5	11.1	553	9	US-10-012-896-113
3	303.5	11.1	553	9	US-09-895-793-113
4	303.5	11.1	553	9	US-09-895-814-113
5	303.5	11.1	553	10	US-09-745-288-101
6	303.5	11.1	553	10	US-09-838-785-2
7	303.5	11.1	553	10	US-09-759-143-113
8	303.5	11.1	553	10	US-09-780-669-113
9	303.5	11.1	553	10	US-09-030-606-113
10	303.5	11.1	553	10	US-09-822-827-113
11	303.5	11.1	553	10	US-09-115-453-113
12	303.5	11.1	1079	9	US-09-895-793-947
13	303.5	11.1	1079	10	US-09-822-827-947
14	234	8.6	359	9	US-09-895-793-974
15	234	8.6	359	10	US-09-822-827-974
16	217.5	8.0	371	9	US-10-012-896-708
17	217.5	8.0	371	9	US-09-895-793-708
18	217.5	8.0	371	9	US-09-895-814-708
19	217.5	8.0	371	10	US-09-759-143-708

371	10	US-09-780-669-708	Sequence 708, App
371	10	US-09-822-827-708	Sequence 708, App
400	9	US-10-012-896-852	Sequence 852, App
400	9	US-09-895-793-852	Sequence 852, App
400	9	US-09-895-814-852	Sequence 852, App
400	10	US-09-759-143-852	Sequence 852, App
400	10	US-09-780-669-852	Sequence 852, App
400	10	US-09-822-827-852	Sequence 852, App
469	10	US-09-925-302-627	Sequence 627, App
70	10	US-09-734-569-124	Sequence 124, App
490	10	US-09-739-457-6	Sequence 6, Appli
490	10	US-09-739-457-7	Sequence 7, Appli
490	10	US-09-739-457-3	Sequence 3, Appli
454	10	US-09-729-674-140	Sequence 140, App
454	10	US-09-739-457-4	Sequence 4, Appli
618	10	US-09-995-007-2	Sequence 2, Appli
460	10	US-09-741-669-464	Sequence 464, App
460	9	US-09-738-626-6145	Sequence 6145, Ap
413	10	US-09-957-664-4	Sequence 4, Appli
448	9	US-09-738-626-6795	Sequence 6795, Ap
504	9	US-09-738-626-7001	Sequence 7001, Ap
497	9	US-09-738-626-6734	Sequence 6734, Ap
1157	9	US-09-738-626-3713	Sequence 3713, Ap
431	10	US-09-815-242-5145	Sequence 5145, Ap
551	9	US-10-216-355-4	Sequence 4, Appli
452	9	US-09-738-626-6848	Sequence 6848, Ap
571	10	US-09-729-674-130	Sequence 130, App
513	10	US-09-815-242-11823	Sequence 11823, A
842	10	US-09-815-242-13718	Sequence 13718, A
444	9	US-09-738-626-6577	Sequence 6577, Ap
457	9	US-10-195-158-6	Sequence 6, Appli
459	9	US-09-738-626-6998	Sequence 6998, Ap
481	9	US-09-738-626-6452	Sequence 6452, Ap
474	9	US-09-828-173-5	Sequence 5, Appli
474	10	US-09-887-038-5	Sequence 5, Appli
424	9	US-09-738-626-3641	Sequence 3641, Ap
528	9	US-10-024-623-5	Sequence 5, Appli
363	9	US-09-802-640-40	Sequence 40, Appl
491	9	US-09-738-626-3703	Sequence 3703, Ap
563	9	US-09-981-947A-2	Sequence 2, Appli
411	9	US-10-084-205-22	Sequence 22, Appl
451	10	US-09-925-637-22	Sequence 22, Appl
466	10	US-09-815-242-13831	Sequence 13831, A
483	9	US-09-738-626-3588	Sequence 3588, Ap
486	9	US-10-051-909-10	Sequence 10, Appl
486	9	US-10-051-902-10	Sequence 10, Appl
727	9	US-09-815-923-10	Sequence 10, Appl
492	9	US-09-978-295A-7	Sequence 7, Appli
492	9	US-09-978-697-7	Sequence 7, Appli
492	9	US-09-978-192A-7	Sequence 7, Appli
492	9	US-09-999-832A-7	Sequence 7, Appli
492	9	US-09-978-189-7	Sequence 7, Appli
492	9	US-09-978-608A-7	Sequence 7, Appli
492	9	US-09-978-191A-7	Sequence 7, Appli
492	9	US-09-978-403A-7	Sequence 7, Appli
492	9	US-09-978-564A-7	Sequence 7, Appli
492	9	US-09-978-585A-7	Sequence 7, Appli
492	9	US-10-017-081A-7	Sequence 7, Appli
492	9	US-09-978-824-7	Sequence 7, Appli
492	9	US-09-981-915A-7	Sequence 7, Appli
492	9	US-09-999-833A-7	Sequence 7, Appli
492	9	US-10-167-749-7	Sequence 7, Appli
492	9	US-09-918-585A-7	Sequence 7, Appli
492	9	US-09-978-423A-7	Sequence 7, Appli
492	9	US-10-013-921A-7	Sequence 7, Appli
551	9	US-09-738-626-4431	Sequence 4431, Ap
284	10	US-09-867-550-1872	Sequence 1872, Ap
53	10	US-09-734-569-122	Sequence 122, Ap
359	9	US-09-738-626-4539	Sequence 4539, Ap
661	9	US-09-738-626-3714	Sequence 3714, Ap
671	10	US-09-843-598-5	Sequence 5, Appli
671	10	US-09-843-598-7	Sequence 7, Appli
2502	10	US-09-772-316-1	Sequence 1, Appli

93 96.5 3.5 453 9 US-09-738-626-4026 Sequence 4026, Ap  
94 96 3.5 539 9 US-10-051-902-26 Sequence 26, Appl  
95 96 3.5 539 9 US-10-051-909-26 Sequence 26, Appl  
96 95.5 3.5 472 10 US-09-822-863-4 Sequence 4, Appl  
97 95.5 3.5 834 10 US-09-815-242-10076 Sequence 10076, A  
98 95.5 3.5 933 10 US-09-815-242-11817 Sequence 11817, A  
99 94.5 3.5 488 9 US-10-170-528-5 Sequence 5, Appl  
100 94.5 3.5 488 9 US-10-062-960B-4 Sequence 4, Appl

## ALIGNMENTS

## RESULT 1

US-09-232-880-113  
Sequence 113, Application US/09232880  
Publication No. US20020182596A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer Lynn  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF  
TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE  
FILE REFERENCE: 210121.428C6  
CURRENT APPLICATION NUMBER: US/09/232.880  
CURRENT FILING DATE: 1999-01-15  
NUMBER OF SEQ ID NOS: 338  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-232-880-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;  
Best Local Similarity 25.3%; Pred. No. 5.4e-18;  
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGWAGVQYQGWALQSLTTPYVOTGLSHALTSPFMMLCGPIAGLVQPLVGLYSD 92  
Db 17 QLLLVNLTFFGLEVCIAAGITVPPLLEVGVEEFMTVMVLGIGVGLVCVPLIGSASD 76  
Qy 93 RCTSRWRRRPFILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPWRHAAIVYVL 152  
Db 77 HWRGRYGRRRPFIFWALSIGLISLFLIPRAGWLGLL-----CP---DPRPLELALLIL 127  
Qy 153 GFWLLDFSNNTVQGPARAMADLC---DHHGSPSAANSIFCSWMLGNLTIGYSSGSTNNWH 209  
Db 128 GVLGLDFCQVCFTPLEALLSDFLRDPDH--CRQAYSVAFAWISLGGCGLYLLPAID--- 182  
Qy 210 KW-----FPFLKTSACCEACANLKGAFIVAVVFLVCLTTLTIFAKEVPPYRANENLPTTK 264  
Db 183 -WDTSAAPYLGTQEEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231  
Qy 265 AGGEVETEPTGPLAVLKGFKDL-----PPGMPSVLLVLTWISWPPFIY 310  
Db 232 LSAPSLSPHCCPCRARLAFRNGLALLPRLHQLCCRPRTLRLFAELCSWMLMFTFLF 291  
Qy 311 DTDWGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPCMKRVG 367  
Db 292 YTFDVEGLYQGVPAEPGTEAR--RHYDEGVRMGSGLGFLQCAISLVFSLVMDRLVQRFG 350  
Qy 368 PRVWVTSNFMVCMVMAAATALISFWSLRDYGIVODATITANASIKAVCLVFAFLGVPPLA 427  
Db 351 TRAVYLASVAAPVPAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400  
Qy 428 ILY 430  
Db 401 SLV 403

## RESULT 2

US-10-012-896-113

Sequence 113, Application US/10012896  
Publication No. US20020183251A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqiu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepner, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Wantanabe, Yoshihiro  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C27  
CURRENT APPLICATION NUMBER: US/10/012.896  
CURRENT FILING DATE: 2001-12-10  
NUMBER OF SEQ ID NOS: 1011  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-896-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;  
Best Local Similarity 25.3%; Pred. No. 5.4e-18;  
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGWAGVQYQGWALQSLTTPYVOTGLSHALTSPFMMLCGPIAGLVQPLVGLYSD 92  
Db 17 QLLLVNLTFFGLEVCIAAGITVPPLLEVGVEEFMTVMVLGIGVGLVCVPLIGSASD 76  
Qy 93 RCTSRWRRRPFILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPWRHAAIVYVL 152  
Db 77 HWRGRYGRRRPFIFWALSIGLISLFLIPRAGWLGLL-----CP---DPRPLELALLIL 127  
Qy 153 GFWLLDFSNNTVQGPARAMADLC---DHHGSPSAANSIFCSWMLGNLTIGYSSGSTNNWH 209  
Db 128 GVLGLDFCQVCFTPLEALLSDFLRDPDH--CRQAYSVAFAWISLGGCGLYLLPAID--- 182  
Qy 210 KW-----FPFLKTSACCEACANLKGAFIVAVVFLVCLTTLTIFAKEVPPYRANENLPTTK 264  
Db 183 -WDTSAAPYLGTQEEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231  
Qy 265 AGGEVETEPTGPLAVLKGFKDL-----PPGMPSVLLVLTWISWPPFIY 310  
Db 232 LSAPSLSPHCCPCRARLAFRNGLALLPRLHQLCCRPRTLRLFAELCSWMLMFTFLF 291  
Qy 311 DTDWGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPCMKRVG 367  
Db 292 YTFDVEGLYQGVPAEPGTEAR--RHYDEGVRMGSGLGFLQCAISLVFSLVMDRLVQRFG 350  
Qy 368 PRVWVTSNFMVCMVMAAATALISFWSLRDYGIVODATITANASIKAVCLVFAFLGVPPLA 427  
Db 351 TRAVYLASVAAPVPAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400  
Qy 428 ILY 430

```

Db 401 SLY 403
||
RESULT 3
US-09-895-793-113
; Sequence 113, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-895-793-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGWAGVQVGMALQSLTPYVOTLGLSHALTSFWMWLCGPIAGLVVQPLVGLYS 92
Db 17 QLLLVNLLTFGLVCLAAAGITVPPLLEVGVEEKFMTVLGIGVPLGLVCVPLLSASD 76
Qy 93 RCTSRGRRRRFFILTCGMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRWHAIVYVL 152
Db 77 HWRGRRRRFFIWLALSIGILLSLFLIPRAGWLAGLL-----CP---DPRPLELALL 127
Qy 153 GFWLLDFSNNTVQGPARAMMADLC---DHHGPSAANSIFCSWMALGNLTIGYSSGSTNNWH 209
Db 128 GVGLLDFCGQVCFPLEALLSDLFDPDPH--CRQAYSVYAFMISLGGCLGLVLLPAID 182
Qy 210 KW-----PPFLKTSACCACANLKGAFVAVVFLVCLTVTLIPAKEVPYRANENLPTK 264
Db 183 -WDTSAAPYLGTOEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
Qy 265 AGGEVEETEPTGLAVLKGKDL-----PGMPSVLLVTAITWLSNFPETLY 310
Db 232 LSAPSLPHCCPCRRARLAFLRNGLALLPRHLQCCMPRTLRRLFVAELCSWMLMTFTLF 291
Qy 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVVRGAFGLLLNSVILGFSFLLIEPMCKRVG 367
Db 292 YTFVGEGLYGVRAEPGTAR-RHYDEGVMSGLGLFLQALISLVSFLVMDLVRQFG 350
Qy 368 PRVWVTSNFVCMVAMAATALISFWSLRDHYGVQDAITANASIKAVCLVLFAELGVPLA 427

Db 351 TRAVYLASVAAPVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
||
Qy 428 ILY 430
||
Db 401 SLY 403
||
RESULT 4
US-09-895-814-113
; Sequence 113, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-895-814-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGWAGVQVGMALQSLTPYVOTLGLSHALTSFWMWLCGPIAGLVVQPLVGLYS 92
Db 17 QLLLVNLLTFGLVCLAAAGITVPPLLEVGVEEKFMTVLGIGVPLGLVCVPLLSASD 76
Qy 93 RCTSRGRRRRFFILTCGMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRWHAIVYVL 152
Db 77 HWRGRRRRFFIWLALSIGILLSLFLIPRAGWLAGLL-----CP---DPRPLELALL 127
Qy 153 GFWLLDFSNNTVQGPARAMMADLC---DHHGPSAANSIFCSWMALGNLTIGYSSGSTNNWH 209
Db 128 GVGLLDFCGQVCFPLEALLSDLFDPDPH--CRQAYSVYAFMISLGGCLGLVLLPAID 182
Qy 210 KW-----PPFLKTSACCACANLKGAFVAVVFLVCLTVTLIPAKEVPYRANENLPTK 264
Db 183 -WDTSAAPYLGTOEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
Qy 265 AGGEVEETEPTGLAVLKGKDL-----PGMPSVLLVTAITWLSNFPETLY 310
Db 232 LSAPSLPHCCPCRRARLAFLRNGLALLPRHLQCCMPRTLRRLFVAELCSWMLMTFTLF 291
Qy 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVVRGAFGLLLNSVILGFSFLLIEPMCKRVG 367

```

```

Db 292 YTDVFGEGLYQGVPRAPGTEAR-RHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 350
QY 368 PRVWVWTSNFMVVCVMAATALISFWSLDYHGYVDATANASIKAVCLVLFPAFLGVPLA 427
Db 351 TRAVYLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILY 430
Db 401 SLY 403

RESULT 5
US-09-745-288-101
; Sequence 101, Application US/09745288
; Patent No. US20010018058A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.446D1
; CURRENT APPLICATION NUMBER: US/09/745,288
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-745-288-101

Query Match 11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGVAGVQYGVGWAQLSLTTPYVOTLGLSHALTSMFWMLCGPIAGLVQPLVGLYSD 92
Db 17 QLLLVNLTFFGLEVCCLAAGITVPPLLLEVGVEEKFMTWLVGIGPVGLVCVPLIGSASD 76
QY 93 RCTSRWRRRRPILTCGMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRWHAIVVVL 152
Db 77 HWRGRRRRPFIWALSGLISLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
QY 153 GFWLLDFSNTVQGPARAMADLC---DHHGPSAANSIFCSMMALGNILGYSSGSTNNWH 209
Db 128 GVLGLDFCQGVCTPLEALLSLDFRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLGAFLVAVVFLVCLTTLIFAKEVPPYRANENLPTTK 264
Db 183 -NDSALAPYLCTQEEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPT-PAEG 231
QY 265 AGGEVETEPTGLAVLKGFKDL-----PGMPSVLLVTAITWLSWFPFLY 310
Db 232 LSAPLSPHCCPCRARLAFRNIGALLPRLHQLCCRMPTLRLFLVAELCSMMALMTFLF 291
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGAFLGILLSVILGFSSFLIEPMCKRVG 367
Db 292 YTDVFGEGLYQGVPRAPGTEAR-RHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 350
QY 368 PRVWVWTSNFMVVCVMAATALISFWSLDYHGYVDATANASIKAVCLVLFPAFLGVPLA 427
Db 351 TRAVYLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILY 430
Db 401 SLY 403

RESULT 6
US-09-838-785-2
; Sequence 2, Application US/09838785
```

```

; Patent No. US20020009455A1
; GENERAL INFORMATION:
; APPLICANT: Lau, Ted
; APPLICANT: Lin, Rick
; APPLICANT: Parkes, Debbie
; APPLICANT: Parry, Gordon
; APPLICANT: Schneider, Douglas
; APPLICANT: Steinbrecher, Renate
; APPLICANT: Van Heuit, Pam T
; APPLICANT: Wu, John
; TITLE OF INVENTION: DNA Encoding a No. US20020009455A1el PROST 03
; FILE REFERENCE: 51831AUSM1
; CURRENT APPLICATION NUMBER: US/09/838,785
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/200,065
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-838-785-2

Query Match 11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGVAGVQYGVGWAQLSLTTPYVOTLGLSHALTSMFWMLCGPIAGLVQPLVGLYSD 92
Db 17 QLLLVNLTFFGLEVCCLAAGITVPPLLLEVGVEEKFMTWLVGIGPVGLVCVPLIGSASD 76
QY 93 RCTSRWRRRRPILTCGMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRWHAIVVVL 152
Db 77 HWRGRRRRPFIWALSGLISLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
QY 153 GFWLLDFSNTVQGPARAMADLC---DHHGPSAANSIFCSMMALGNILGYSSGSTNNWH 209
Db 128 GVLGLDFCQGVCTPLEALLSLDFRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLGAFLVAVVFLVCLTTLIFAKEVPPYRANENLPTTK 264
Db 183 -WDSALAPYLCTQEEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPT-PAEG 231
QY 265 AGGEVETEPTGLAVLKGFKDL-----PGMPSVLLVTAITWLSWFPFLY 310
Db 232 LSAPLSPHCCPCRARLAFRNIGALLPRLHQLCCRMPTLRLFLVAELCSMMALMTFLF 291
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGAFLGILLSVILGFSSFLIEPMCKRVG 367
Db 292 YTDVFGEGLYQGVPRAPGTEAR-RHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 350
QY 368 PRVWVWTSNFMVVCVMAATALISFWSLDYHGYVDATANASIKAVCLVLFPAFLGVPLA 427
Db 351 TRAVYLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILY 430
Db 401 SLY 403

RESULT 7
US-09-759-143-113
; Sequence 113, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
```



```

; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759.143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-759-143-113

Query Match      11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGVAGGVQVGMALQLSLTPYVOTGLSHALTSMFWMLCGPAGLVQPLVGLYSD 92
Db 17 QLLVNLTLFTGLEVCCLAAGITVPPLLEVEEKFMTVLGIGVGLVCVPLLGSSASD 76
Qy 93 RCTSWGRRRPFILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVYVL 152
Db 77 HWRGRRRRPFIFWALSGLLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127
Qy 153 GFWLDFSNNTVQGPARAMMADLC---DHGSPSAANSIFCSMMALGNTILGYSSGSTNNWH 209
Db 128 VGLLDFCGQCVFTPLEALLSDFRDPDH--CRQAYSVYAFMISLGGCLGYLLPAID--- 182
Qy 210 KW-----PFELKTSACCACANLKGAFLVAVVFLVCLTITLIFAKEVPYRANENLPTK 264
Db 183 -WDTSAALPYLGTQEEC-----LFG--LTLIFLT-CVAATLLVAEEAALGPTG--PAEG 231
Qy 265 AGGEVETEPTGLAVLKGFKDL-----PPGMPSVLLVTAITWLSWFPFITY 310
Db 232 LSAPSLPHCCPCRARLAFRNLGALLPRHLQCCRMPTLRLFLVAELCSNMALMTFTLF 291
Qy 311 DTDWMGREIYHGDPK---GSNAQISAFNEGVRVGAFGLLLNSVILGFSFSLIEPMCKRVG 367
Db 292 YTDVFGEGLYQGVPRAEPTGTEAR-RHYDEGVRMGSLGLFLOCAISLVSLVMDRLVQRF 350
Qy 368 PRVWVVTGNFVWCVVAMAATALISFWSLDRYHGYVODAITANASIKAVCLVLPFAFLGVPLA 427
Db 351 TRAVYLVASVAAPPPAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
Qy 428 ILY 430
Db 401 SLY 403

RESULT 8
US-09-780-669-113
; Sequence 113, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
```

```

; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780.669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-780-669-113

Query Match      11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGVAGGVQVGMALQLSLTPYVOTGLSHALTSMFWMLCGPAGLVQPLVGLYSD 92
Db 17 QLLVNLTLFTGLEVCCLAAGITVPPLLEVEEKFMTVLGIGVGLVCVPLLGSSASD 76
Qy 93 RCTSWGRRRPFILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVYVL 152
Db 77 HWRGRRRRPFIFWALSGLLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127
Qy 153 GFWLDFSNNTVQGPARAMMADLC---DHGSPSAANSIFCSMMALGNTILGYSSGSTNNWH 209
Db 128 VGLLDFCGQCVFTPLEALLSDFRDPDH--CRQAYSVYAFMISLGGCLGYLLPAID--- 182
Qy 210 KW-----PFELKTSACCACANLKGAFLVAVVFLVCLTITLIFAKEVPYRANENLPTK 264
Db 183 -WDTSAALPYLGTQEEC-----LFG--LTLIFLT-CVAATLLVAEEAALGPTG--PAEG 231
Qy 265 AGGEVETEPTGLAVLKGFKDL-----PPGMPSVLLVTAITWLSWFPFITY 310
Db 232 LSAPSLPHCCPCRARLAFRNLGALLPRHLQCCRMPTLRLFLVAELCSNMALMTFTLF 291
Qy 311 DTDWMGREIYHGDPK---GSNAQISAFNEGVRVGAFGLLLNSVILGFSFSLIEPMCKRVG 367
Db 292 YTDVFGEGLYQGVPRAEPTGTEAR-RHYDEGVRMGSLGLFLOCAISLVSLVMDRLVQRF 350
Qy 368 PRVWVVTGNFVWCVVAMAATALISFWSLDRYHGYVODAITANASIKAVCLVLPFAFLGVPLA 427
Db 351 TRAVYLVASVAAPPPAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
Qy 428 ILY 430
Db 401 SLY 403

RESULT 9
US-09-030-606-113
; Sequence 113, Application US/09030606
; Patent No. US20020081580A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHOD
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
```





Db 654 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAFMISLGGCGLYLLPAID--- 708  
QY 210 KW-----PPFLKTSACCEACANLGAFLVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264  
Db 709 -WDTSAAPYLGTOEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 757  
QY 265 AGGEVETEPTGFLAVLKGFKDL-----PGMPSVLLVTITWLSWPPFIY 310  
Db 758 LSAPSLPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRRLFVAELCSNMALMTFTLF 817  
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRGAFGLLLNSVILGFSFLIEPMCRKVG 367  
Db 818 YTDVFGEGLYQGVPRAPGTEAR-RHYDEGVRMGSGLGLFLOCAISLVFSLVMDRLVQRFG 876  
QY 368 PRVWVWTSFMVCMVAMATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427  
Db 877 TRAVILASVAAPFVAAGATCL-----SHSVAVVTSAAALGTFTSALQILPYTLA 926  
QY 428 ILY 430  
Db 927 SLY 929

## RESULT 14

US-09-895-793-974  
; Sequence 974, Application US/09895793  
; Publication No. US20020192763A1

## GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Hural, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Vinals de Bassols, Carlota  
; APPLICANT: Foy, Teresa  
; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C2

; CURRENT APPLICATION NUMBER: US/09/895,793

; CURRENT FILING DATE: 2001-06-29

; NUMBER OF SEQ ID NOS: 982

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 974

; LENGTH: 359

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-895-793-974

Query Match 8.6%; Score 234; DB 9; Length 359;  
Best Local Similarity 23.9%; Pred. No. 2.9e-12;  
Matches 91; Conservative 69; Mismatches 147; Indels 74; Gaps 16;

QY 33 RLILAGWAGGVQYQWALQSLTTPYVQTLGLSHALTSMWLCGPIAGLVQPLVGLYSD 92

Db 17 QLLLVNLTFFGLEVCLAAGITVPPLLLEVGVEEKFMTWLVGIGVPLVLCVPLIGSASD 76

QY 93 RCTSRGRRRPILTGCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVL 152

Db 77 HWRGRYGRRRPFITWALSGLILLSFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127  
QY 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSNMALGNITLGYSSGSTNNWH 209  
Db 128 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAFMISLGGCGLYLLPAID--- 182  
QY 210 KW-----PPFLKTSACCEACANLGAFLVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264  
Db 183 -WDTSAAPYLGTOEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231  
QY 265 AGGEVETEPTGFLAVLKGFKDL-----PGMPSVLLVTITWLSWPPFIY 310  
Db 232 LSAPSLPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRRLFVAELCSNMALMTFTLF 291  
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRGAFGLLLNSVILGFSFLIEPMCRKVG 367  
Db 292 YTDVFGEGLYQGVPRAPGTEAR-RHYDEGVRMGSGLGLFLOCAISLVFSLVMDRLVQRFG 876  
QY 368 PRVWVWTSFMVCMVAMATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427  
Db 877 TRAVILASVAAPFVAAGATCL-----SHSVAVVTSAAALGTFTSALQILPYTLA 926  
QY 428 ILY 430  
Db 927 SLY 929

## RESULT 15

US-09-822-827-974  
; Sequence 974, Application US/09822827  
; Patent No. US20020081680A1

## GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C1  
; CURRENT APPLICATION NUMBER: US/09/822,827  
; CURRENT FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 974  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-822-827-974

Query Match 8.6%; Score 234; DB 10; Length 359;  
Best Local Similarity 23.9%; Pred. No. 2.9e-12;  
Matches 91; Conservative 69; Mismatches 147; Indels 74; Gaps 16;

QY 33 RLILAGWAGGVQYQWALQSLTTPYVQTLGLSHALTSMWLCGPIAGLVQPLVGLYSD 92

Db 17 QLLLVNLTFFGLEVCLAAGITVPPLLLEVGVEEKFMTWLVGIGVPLVLCVPLIGSASD 76

QY 93 RCTSRGRRRPILTGCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVL 152

Db 77 HWRGRYGRRRPFITWALSGLILLSFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127

QY 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSNMALGNITLGYSSGSTNNWH 209

Db 128 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAFMISLGGCGLYLLPAID--- 182

QY 210 KW-----PPFLKTSACCEACANLGAFLVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264

Db 183 -WDTSAAPYLGTOEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231

QY 265 AGGEVETEPTGFLAVLKGFKDL-----PGMPSVLLVTITWLSWPPFIY 310

Db 232 LSAPSLPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRRLFVAELCSNMALMTFTLF 291

QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRGAFGLLLNSVILGFSFLIEPMCRKVG 367

Db 292 YTDVFGEGLYQGVPRAPGTEAR-RHYDEGVRMGSGLGLFLOCAISLVFSLVMDRLVQRFG 876

QY 360 EPMCRKVGPRVWVWTSFMVCMVAMATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427

Db 342 ---CAQVHPQKV---TKFMLC 356

## RESULT 16

US-10-012-896-708  
; Sequence 708, Application US/10012896  
; Publication No. US20020183251A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Devin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedrick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Hural, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Vinals de Bassols, Carlota  
; APPLICANT: Foy, Teresa  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Wantanabe, Yoshihiro  
; APPLICANT: Meagher, Madeleine Joy  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.427C27  
; CURRENT APPLICATION NUMBER: US/10/012.896  
; CURRENT FILING DATE: 2001-12-10  
; NUMBER OF SEQ ID NOS: 1011  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 708  
; LENGTH: 371  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-012-896-708

Query Match 8.0%; Score 217.5; DB 9; Length 371;  
Best Local Similarity 26.7%; Pred. No. 8e-11;  
Matches 99; Conservative 52; Mismatches 147; Indels 73; Gaps 19;

Qy 55 LTPYVOTLGLSHAL---TSFMWLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFILTGCM 111  
Db 10 LVPLPLALYLSQPLTHTTSLLAGIGPVLCVPLGSLGSDHWRGRRRPFIWALSIG 69

Qy 112 ICVAVIVVGFSSDICAALGDTKEHCSLYHGPRWHAIVVYVGLFWLLDFSNNTVQGP 171  
Db 70 ILLSLFLPRAGWLAGLL-----CP---DPRPLEALLILGVLGLDFCGQVCFPTPLEAL 120

Qy 172 MADLC---DHHGPSAANSIFCSWMAIGNILGYSSGSTNNHWK-----FPFLKTSACCEA 223  
Db 121 LSDLRDPDH--CROQSVYAFMISLGGCLGYLLPAID----WDTSALAPYLGTOEC--- 172

Qy 224 CANLKGAFLVAVVFLVCLTTLTIFAKEVPYRANENLPTTKAGGEVETEPTGPLAVLKGF 283  
Db 173 ---LFG--LLTLIFLT-CVAATLLVAEEAALGPTE--PAEGLSAPSLSPHCCPCRLAF 224

Qy 284 KDL-----PPGMPSVLLVTAITWLSWFFPILYDTDMGRIYHGDPK---G 326  
Db 225 RNLGALLPRLHQLCCMRPTRLRLEVAELCSWMLMTLTFTYDFVGEGLYQGVPRABPG 284

Qy 327 SNAQISAFNEGVRVCA---FGLLLNSVLTGFSFLIEPM-----CRKVGPRVV---- 371  
Db 285 TEAR--RHYDEGKALAASRGWCGSRPPTTLGAVSGLV-PLHPGPDFSVRKVGMDPICI 342

Qy 372 --WVTSNFMVC 380

Db 343 FSWWN-NISAC 352

## RESULT 17

US-09-895-793-708  
; Sequence 708, Application US/09895793  
; Publication No. US20020192763A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Devin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedrick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Hural, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Vinals de Bassols, Carlota  
; APPLICANT: Foy, Teresa  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C2  
; CURRENT APPLICATION NUMBER: US/09/895,793  
; CURRENT FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 708  
; LENGTH: 371  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-895-793-708

Query Match 8.0%; Score 217.5; DB 9; Length 371;  
Best Local Similarity 26.7%; Pred. No. 8e-11;  
Matches 99; Conservative 52; Mismatches 147; Indels 73; Gaps 19;

Qy 55 LTPYVOTLGLSHAL---TSFMWLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFILTGCM 111  
Db 10 LVPLPLALYLSQPLTHTTSLLAGIGPVLCVPLGSLGSDHWRGRRRPFIWALSIG 69

Qy 112 ICVAVIVVGFSSDICAALGDTKEHCSLYHGPRWHAIVVYVGLFWLLDFSNNTVQGP 171  
Db 70 ILLSLFLPRAGWLAGLL-----CP---DPRPLEALLILGVLGLDFCGQVCFPTPLEAL 120

Qy 172 MADLC---DHHGPSAANSIFCSWMAIGNILGYSSGSTNNHWK-----FPFLKTSACCEA 223  
Db 121 LSDLRDPDH--CROQSVYAFMISLGGCLGYLLPAID----WDTSALAPYLGTOEC--- 172

Qy 224 CANLKGAFLVAVVFLVCLTTLTIFAKEVPYRANENLPTTKAGGEVETEPTGPLAVLKGF 283  
Db 173 ---LFG--LLTLIFLT-CVAATLLVAEEAALGPTE--PAEGLSAPSLSPHCCPCRLAF 224

Qy 284 KDL-----PPGMPSVLLVTAITWLSWFFPILYDTDMGRIYHGDPK---G 326  
Db 225 RNLGALLPRLHQLCCMRPTRLRLEVAELCSWMLMTLTFTYDFVGEGLYQGVPRABPG 284

Qy 327 SNAQISAFNEGVRVCA---FGLLLNSVLTGFSFLIEPM-----CRKVGPRVV---- 371  
Db 285 TEAR--RHYDEGKALAASRGWCGSRPPTTLGAVSGLV-PLHPGPDFSVRKVGMDPICI 342

Qy 372 --WVTSNFMVC 380





APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
APPLICANT: Wantanabe, Yoshihiro  
APPLICANT: Meagher, Madeleine Joy  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C27  
CURRENT APPLICATION NUMBER: US/10/012,896  
CURRENT FILING DATE: 2001-12-10  
NUMBER OF SEQ ID NOS: 1011  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 852  
LENGTH: 400  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-896-852

Query Match 7.0%; Score 191.5; DB 9; Length 400;  
Best Local Similarity 24.7%; Pred. No. 1.5e-08;  
Matches 71; Conservative 49; Mismatches 120; Indels 47; Gaps 11;

QY 52 LSLTPYVOTLGLSHALTSMMLCGPIAGLVQPLVGLYSDRCTSRWGRRRPFILTCML 111  
DB 138 ITYVPLLEVGVEEFMTWLGIGVLGVCVPLIGSASDHWRGRRRPFILWALSIG 197  
QY 112 ICVAVIVGSSDGAALGDTKEHCSLYHGPWHAIAVYVGLFWLLDFSNNTVQGPARAM 171  
DB 198 ILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLILGVLLDFCGQVCFTPLEAL 248  
QY 172 MADLC---DHHGPSAANSIFCSWMAIGNILGYSSGSTNNHKKW-----FPFLKTSACCEA 223  
DB 249 LSDLFRDPDH--CROAYSVYAFMISLGGCLGYLLPAID----WDTSAAPYLGTOEEC-- 300  
QY 224 CANLKAGFLVAVVFLVCLTTLIFAKEVPYRANENLPTTKAGGEVEETEPTGLAVLKG 283  
DB 301 ---LFG--LLTLIFLT-CVAATLIVAEAAALGPT--PAEGLSAPLSLPHCCPCRARLAF 352  
QY 284 KDL-----PPGMSVLLVTAITWLSWFPFPLYDWDWG 316  
DB 353 RNUGALLPRLHQLCCMPRTLRLFLVAELCSWMLMTFTFLFYDFVG 399

RESULT 23  
US-09-793-793-852  
Sequence 852, Application US/09895793  
Publication No. US20020192763A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqiu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.534C2

CURRENT APPLICATION NUMBER: US/09/895,793  
CURRENT FILING DATE: 2001-06-29  
NUMBER OF SEQ ID NOS: 982  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 852  
LENGTH: 400  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-895-793-852

Query Match 7.0%; Score 191.5; DB 9; Length 400;  
Best Local Similarity 24.7%; Pred. No. 1.5e-08;  
Matches 71; Conservative 49; Mismatches 120; Indels 47; Gaps 11;

QY 52 LSLTPYVOTLGLSHALTSMMLCGPIAGLVQPLVGLYSDRCTSRWGRRRPFILTCML 111  
DB 138 ITYVPLLEVGVEEFMTWLGIGVLGVCVPLIGSASDHWRGRRRPFILWALSIG 197  
QY 112 ICVAVIVGSSDGAALGDTKEHCSLYHGPWHAIAVYVGLFWLLDFSNNTVQGPARAM 171  
DB 198 ILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLILGVLLDFCGQVCFTPLEAL 248  
QY 172 MADLC---DHHGPSAANSIFCSWMAIGNILGYSSGSTNNHKKW-----FPFLKTSACCEA 223  
DB 249 LSDLFRDPDH--CROAYSVYAFMISLGGCLGYLLPAID----WDTSAAPYLGTOEEC-- 300  
QY 224 CANLKAGFLVAVVFLVCLTTLIFAKEVPYRANENLPTTKAGGEVEETEPTGLAVLKG 283  
DB 301 ---LFG--LLTLIFLT-CVAATLIVAEAAALGPT--PAEGLSAPLSLPHCCPCRARLAF 352  
QY 284 KDL-----PPGMSVLLVTAITWLSWFPFPLYDWDWG 316  
DB 353 RNUGALLPRLHQLCCMPRTLRLFLVAELCSWMLMTFTFLFYDFVG 399

RESULT 24  
US-09-895-814-852  
Sequence 852, Application US/09895814  
Publication No. US20020193296A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqiu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C26  
CURRENT APPLICATION NUMBER: US/09/895,814  
CURRENT FILING DATE: 2001-06-29  
NUMBER OF SEQ ID NOS: 990  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 852  
LENGTH: 400  
TYPE: PRT  
ORGANISM: Homo sapiens







GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:30:00 ; Search time 22 Seconds  
(without alignments)  
2267.898 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSAAGVRGAA.....VFLPKISKRQFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR\_73:\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2317.5	84.9	537	2 T02982	probable sucrose t
2	1523	55.8	594	2 G84441	probable sucrose-p
3	1144	41.9	533	2 S43142	sucrose transport
4	1126	41.2	501	2 T14339	sucrose-proton tra
5	1116	40.9	510	2 S38557	sucrose transport
6	1112.5	40.7	515	2 T14340	sucrose transport
7	1110	40.6	523	2 T12198	sucrose-proton tra
8	1101.5	40.3	492	2 A84520	probable sucrose-p
9	1100.5	40.3	513	2 S38197	sucrose transport
10	1095.5	40.1	491	2 F96591	probable sucrose-p
11	1094	40.1	512	2 S38196	sucrose transport
12	1088	39.8	512	2 G86360	probable sucrose-p
13	1081	39.6	525	2 S28052	sucrose transport
14	1079	39.5	516	2 JQ2389	sucrose transport
15	1074	39.3	503	2 S52377	sucrose transport
16	1070	39.2	512	2 F96741	probable sucrose t
17	1063.5	38.9	507	2 S48789	sucrose transport
18	1043.5	38.2	523	2 S51114	sucrose-proton sym
19	996	36.5	474	2 A86234	hypothetical prote
20	883.5	32.4	428	2 S48788	sucrose transport
21	349	12.8	553	2 T38541	probable sucrose c
22	305.5	11.2	541	2 B87532	transporter, proba
23	246	9.0	451	2 E81781	probable integral
24	245	9.0	451	2 A81206	sugar transporter,
25	210.5	7.7	454	2 A75444	hypothetical prote
26	193.5	7.1	452	2 F75217	hypothetical prote
27	162.5	6.0	430	2 E75217	transporter PAB217
28	160	5.9	418	2 B87536	membrane protein,
29	148.5	5.4	490	2 JC5641	sugar transporter

orf3 bah 3'-region  
drug efflux protei  
tetracycline resis  
probable NADH dehy  
Na+/H+ antiporter  
melibiose carrier  
hypothetical prote  
drug transport pro  
probable membrane  
probable efflux pr  
MFS permease [drug  
probable inorganic  
probable nitrate/n  
probable inner mem  
probable membrane  
probable amino aci  
thyroid iodide tra  
glycerol-3-phospha  
probable transport  
hypothetical prote  
sugar,proton sympo  
probable n-hydroxy  
probable melibiose  
probable permease  
carbon starvation  
probable carbon st  
probable carbon st  
hypothetical prote  
probable transport  
probable transport  
probable transport  
hypothetical prote  
potassium uptake p  
hypothetical prote  
probable transcrip  
probable transport  
probable multidrug  
hypothetical 51.0  
hypothetical prote  
hypothetical prote  
citrate-carrier pr  
Pfs enzyme II, ABC  
L-arabinose transp  
xyloside transport  
high-affinity nitr  
probable MFS trans  
permeases homolog  
probable MFS trans  
chloramphenicol re  
probable MFS trans  
probable hexose tr  
sodium iodide symp  
hypothetical prote  
sugar/Na+(H+) slmp  
probable transport  
major facilitator  
metabolite transpo  
bacteriochlorophyl  
major facilitator  
probable MFS trans  
probable transport  
probable transport  
nitrogen fixation  
hypothetical prote  
hypothetical prote  
proline permease  
glucose transport  
probable MFS trans

ALIGNMENTS

RESULT 1

T02982

probable sucrose transport protein - rice

C:Species: Oryza sativa (rice)

C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000

C:Accession: T02982

R:Hitose, T.; Imaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.

Plant Cell Physiol. 38, 1389-1396, 1997

A:Title: cDNA cloning and tissue specific expression of a gene for sucrose transporter

A:Reference number: Z14809; MUID:98182940; PMID:9522469

A:Accession: T02982

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-537 <H1>

A:Cross-references: EMBL:D87819; NID:q2723470; PIDN:BAA24071.1; PID:q2723471

A:Experimental source: cultivar Nipponbare, leaf

C:Genetics:

A:Gene: SUT1

C:Superfamily: common tobacco sucrose transport protein

Query Match 84.9%; Score 2317.5; DB 2; Length 537;

Best Local Similarity 81.2%; Pred. No. 2.8e-176;

Matches 435; Conservative 39; Mismatches 43; Indels 19; Gaps 3;

QY 1 MARGD-----GGQAEASGVGR-----AAAVVDHVAPIISLGRLLILAGVAG 42

DB 1 MARGSGAGGGGGGGGGLSVGVGGGARGGGGEAAATAPIISLGRLLILSGVAG 60

QY 43 GVOYGWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGLYSDRCTSRWGR 102

DB 61 GVOYGWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGLYSDRCTSRWGR 120

QY 103 PRILGCMILICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIVVYLGFWLLDFSN 162

DB 121 PYLTGCVLLICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIVVYLGFWLLDFSN 180

QY 163 TVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHWHKFFPKTSACE 222

DB 181 TVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHWHKFFPKTSACE 240

QY 223 ACANLKGAFVAVVFLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPTGLPLVLKG 282

DB 241 ACANLKGAFVAVVFLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPTGLPLVLKG 299

QY 283 FKDLPPGMPVLLVTAITWLSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVVGA 342

DB 300 FRLNPTGMPVLLVTAITWLSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVVGA 359

QY 343 FGLLNSVILGFSFLIEPCKRKVPRVWVTSNPMVCVMAAATALISFWSLRDYHGVO 402

DB 360 FGLLNSVILGFSFLIEPCKRKVPRVWVTSNPMVCVMAAATALISFWSLRDYHGVO 419

QY 403 DAITANASIKAVCLVFLAFGLVPLAIVSVFPAVTAQALATRGQGLCTGVNLNIVIP 462

DB 420 KAITADKSIKAVCLVFLAFGLVPLAIVSVFPAVTAQALATRGQGLCTGVNLNIVIP 479

QY 463 QVIIALGAGPMDALFGKGNIPAFGVSAPALVGVGVVFLPKISKROFRAVSAGG 518

DB 480 QVIIALGAGPMDALFGKGNIPAFGVSAPALVGVGVVFLPKISKROFRAVSAGG 535

RESULT 2

G84441

probable sucrose-proton symporter [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C:Accession: G84441

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84441

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-594 <STO>

A:Cross-references: GB:AE002093; NID:g3461813; PIDN:AAC32907.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g02860

A:Map position: 2

C:Superfamily: common tobacco sucrose transport protein

Query Match 55.8%; Score 1523; DB 2; Length 594;

Best Local Similarity 53.6%; Pred. No. 3.7e-113;

Matches 288; Conservative 82; Mismatches 117; Indels 50; Gaps 5;

QY 30 SLGRLLIAGVAGVGVQWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGL 89

DB 59 SLVTLVLSCTVAAGVQWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGL 118

QY 90 YSDRCTSRWGRRPFIITGCMILICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIV 149

DB 119 WSDRCTSRWGRRPFIITGCMILICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIV 178

QY 150 YVLGFWLLDFSNNTVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHWH 209

DB 179 FIIGFWLLDLNNTVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHWH 238

QY 210 KWFFPLKTSACCACANLKGAFVAVVFLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPTGL 258

DB 239 EMFPELTSTRACCAACGNLKAFLVAVVFLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPTGL 298

QY 259 -----NLPTTAGG-----EVETEPTGLAV-----LKG 282

DB 299 DDLSKGLHSLKNGTANGIKYVERDTEQFGNSENSEHEDTYYDGGPSVLNLLTS 358

QY 283 FKDLPPGMPVLLVTAITWLSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVVGA 342

DB 359 LRHLPPAMHSLVILNALTWLSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVVGA 418

QY 343 FGLLNSVILGFSFLIEPCKRKVPRVWVTSNPMVCVMAAATALISFWSLRDYHGVO 402

DB 419 LGLLNSVILGFSFLIEPCKRKVPRVWVTSNPMVCVMAAATALISFWSLRDYHGVO 478

QY 403 DAITANASIKAVCLVFLAFGLVPLAIVSVFPAVTAQALATRGQGLCTGVNLNIVIP 462

DB 479 YIMRGNETTRTAIVVIFALLGFLAIVSVFPAVTAQALATRGQGLCTGVNLNIVIP 538

QY 463 QVIIALGAGPMDALFGKGNIPAFGVSAPALVGVGVVFLPKISKROFRAVSAGG 519

DB 539 QMIVSLGAGPMDQLFGGGLPAFVLAFAAGVIALQRLPTLSS-SFK--STGFH 592

RESULT 3

S43142

sucrose transport protein - castor bean

N:Alternate names: sucrose carrier

C:Species: Ricinus communis (castor bean)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-May-2000

C:Accession: S43142

R:Weig, A.; Komor, E.

submitted to the EMBL Data Library, March 1994

A:Description: A sucrose carrier from Ricinus communis.

A:Reference number: S43142

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-533 <WEI>

A:Cross-references: EMBL:Z31561; NID:g468561; PID:g468562

C:Superfamily: common tobacco sucrose transport protein

Query Match

41.9%; Score 1144; DB 2; Length 533;





	Query Match	40.3%;	Score 1100.5;	DB 2;	Length 513;
	Best Local Similarity	42.9%;	Pred. No. 1.1e-79;		
	Matches 222; Conservative	102;	Mismatches 173;	Indels 21;	Gaps 7;
QY	11 ELSAGVRGAAAV-----VDHVAPTSIGRLTLAGVAGGVQVGWALQLSLTTPPYVTGLGS	65			
	:       :	:	:  :  :  :  :  :  :  :  :		
Dd	5 ETEKPTDAAALETQSPEDFQSPLRKIISVASIAGVQFGWALQLSLTTPPVQLLGIP	64			
	:       :	:	:  :  :  :  :  :  :  :  :		
QY	66 HALTSTPMWLCGPITAGLVVQPLVLGYSDRCSTRNGRRRPFILTCGMCLTCVAVIVVGFSSDI	125			
	:         :  :  :  :  :  :  :  :  :  :  :  :  :				
Dd	65 HKWSSLTWLCCPVSGMTVPQIVGFHSDRCRSKFGRRRPFATGAALVAVAVFLIGYAADF	124			
	:         :  :  :  :  :  :  :  :  :  :  :  :  :				

[illegible]

QY	302	LSMFILHIDIDMGRALIHIDRFGSNRQKLSRKNRGRVAVRQGLLRSIVLIDRSGSILRFP	501
Db	295	IAWFPELLFDIDMGRREVFGSDGNSRKKLSLGVQSGAMGLMFSIVILGFNMSLGVSW	354
QY	362	MCRKV-GPRVVVYTSNFMVCVMAATALISFWSLRDYGHVQDAITANASTKACVCLVLEA	420
Db	355	IGRKLGGAKRLWGINVFILAAGLAMTVLVTKFA-EDHRKTAGDLAGFSASVKAGALSLEA	413
QY	421	FLGVPLAILYSVPFAVTAQAAATRGGGQGLCTGVLNISIVIPQVIAALGAGPMDALFKG	480
Db	414	VLGIPLAITESTPFALASIFSSCSGAGQGLSLGVNLNLAIVIPQIVSLGGGPPDLEGGG	473
QY	481	NIPAFGVASAFALVGGVGVFLLPKISKRFRAVSAGG	518
Db	474	NLPAFIVAAIAAASIGVALTVLPSPPPDPAPKATTMGG	511

RESULT 10  
E96691  
probable sucrose-proton symporter T1217.2 [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: E96691  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; D  
ansen, N.P.; Hughes, B.; Huizlar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.

A:Accession: E96691  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-491 >STO>

A; cross-references: GB:AE0051173; NID:g11054577; PIDN:AAG27852.1; GSPDB:GN00141  
 C; Genetics:  
 A; Gene: T12I7.2  
 A; Map position: 1  
 C; Superfamily: common tobacco sucrose transport protein

Query Match	40.18;	Score	1095.5;	DB	2;	Length	491;
Best Local Similarity	43.28;	Pred. No.	2.7e-79;				
Matches	213;	Conservative	101;	Mismatches	158;	Indels	21;
						Gaps	6;





Db 66 WASLIWLCGPTSGMLVQPIVGYHSDRCTSRGRRRPPIVAGAGLVTVAVFLIGYAADIGH 125  
Qy 128 ALGDTKEHCSLYHGRHAAITVYVGLFWLLDFSNVTQGPARAMMADLCHHGFS--AAN 185  
Db 126 SMGB-----QDKPKPTRAIAIFALGFWLLDVANNTLQGPCRAFLADLSAGNAKKTATN 180  
Qy 186 SIFCSMALGNILGYSGSTNNHKKWPFLLKTSACCACANLKGAFVLVAVVFLVLCITVT 245  
Db 181 AFFSFFNAVGNVLYAAGSYRNLYKVPFTTESCDLYCANLKTCCFELSITLLIVTFVS 240  
Qy 246 LIFAKEVPIRANENLPTTKAGGEVETPTGPLAVLKGFKOLPPONPSVLLVTAITWLSWF 305  
Db 241 LCYVYKERPW-----TPEPTADGKASNPFFG-EIFGAFKELKPMWMLLIIVTALNWTAF 294  
Qy 306 PFILYDMDMGREIYHG--DPKGSNAQISAFNEGVRVGAFLGLLNSVLIGFSFLIEPMC 363  
Db 295 PFLLFDMDMGREYVYGGNSDATAAASKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIG 354  
Qy 364 RKV-GPRVWVWTSNFMVCMVAMAATLISFWS---LRDHYGVQDAITANASIKAVCLVLF 419  
Db 355 RKLGAARLWGVNFIILAICLAIVTVTKQAEHRRD-HGAKTGPPGN--VTAGALTIF 411  
Qy 420 AFLGVPLAIIYVFPANTQAATRGQGLCTGVNLISIVIPQVITIALGAGPMDALFGK 479  
Db 412 AILGIPQAITSIPFALASIFSTNSGAGQGLSLGLVNLAIIVVQMVSVGGGPPDELF 471  
Qy 480 GNIPAFGVASAFALVGGVGVFLLP 504  
Db 472 GNIPAFVIGATAAASVGLALTVP 496

RESULT 13  
S28052  
sucrose transport protein - spinach  
N:Alternate names: sucrose carrier protein; sucrose permease  
C:Species: Spinacia oleracea (spinach)  
C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
C:Accession: S28052  
R:Riesmeier, J.W.; Willmitzer, L.; Frommer, W.B.  
EMBO J. 11, 4705-4713, 1992  
A:Title: Isolation and characterization of a sucrose carrier cDNA from spinach by functional complementation  
A:Reference number: S28052; MUID:93099843; PMID:1464305  
A:Accession: S28052  
A:Molecule type: mRNA  
A:Residues: 1-525 <RTE>  
A:Cross-references: EMBL:X67125; NID:g21318; PIDN:CAA47604.1; PID:g21319  
A:Superfamily: common tobacco sucrose transport protein  
C:Keywords: transmembrane protein

Query Match 39.6%; Score 1081; DB 2; Length 525;  
Best Local Similarity 44.9%; Pred. No. 4e-78;  
Matches 227; Conservative 89; Mismatches 166; Indels 24; Gaps 10;

Qy 27 APISLGRILAGVAGVQVGMALQSLTPYVOTLGLSHALTFSFMLCGPIAGLVVQPL 86  
Db 31 AEATLKKLGLVASAAGVQVGMALQSLTPYVQLLGTPHTWAAIYIWLCPISGMIVQPL 90  
Qy 87 VGLYSDRCTSRWGRRRPFIITGCMILCVAVIVGVFSSDIGAALGDTKEHCSLYHGRWHA 146  
Db 91 VGYISDRCTSRGRRRPFIAAGAAALVAVAVGLIGFAADGAASGDPTGNVA---KPR--A 145  
Qy 147 AIVVVLGFWLLDFSNVTQGPARAMMADLC--DHGFSANISIFCSNMALGNILGYSGS 204  
Db 146 IAVFVGVFWLLDVANNTLQGPCRALLADMAAGSQTKRYANAFSFFMALGNIGCYAGS 205  
Qy 205 TNNHKKWPFLLKTSACCACANLKGAFVLVAVVFLVLCITVLIPIAKE-----VPIRANEN 259  
Db 206 YSRLYTVFPFTKTAACDYCANLKSCEFFISITLLIVLTILALSVMVKERQITIDEIQEED 265  
Qy 260 LPT--TKAGGEVETPTGPLAVLKGFKDLPQCMPSVLLVTAITWLSWPFLLYDWDWGRE 318  
Db 266 LKNRNSSGCARLPFFGQL--IGALKOLPKPMLILLVLTALNWTAFWFFLLFDTDWNGKE 323

Qy 319 IYHGDPKGSNAQISAFNEGVRVGAFLGLLNSVLIGFSFLIEPMCKRY-GPRVWVWTSNF 377  
Db 324 VY-----GGTVGEGKLYDQGVHAGALGLMINSVLGVWLSIEGLRVMVGAKRLWGVNI 379  
Qy 378 MVCVAMAATLII--SFWSLRDHYGVQDAI--TANASTKAVCLVLFAPLGVPLAILYSVP 433  
Db 380 ILAVCLAMTLVTKSAEHFRDHHIMSGSVPVPPPPAGVKGALAIFAVLGIPLAITFISIP 439  
Qy 434 FAVTAQAAATRGQGLCTGVNLISIVIPQVITIALGAGPMDALFGKNIPAFGVASAFAL 493  
Db 440 FALASIFASGSGQGLSLGLVNLAIIVVQMFVSVTSGPMDAMFGGGLNLPFAVVGAVAT 499  
Qy 494 VGGVGVFLLPKISKROFRAVSAGGH 519  
Db 500 ASAVLSFTLLPSPPEAKIGSGMGH 525

RESULT 14  
JQ2389  
sucrose transport protein - potato  
C:Species: Solanum tuberosum (potato)  
C>Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 21-Jul-2000  
C:Accession: JQ2389; S40310  
R:Riesmeier, J.W.; Hirner, B.; Frommer, W.B.  
Plant Cell 5, 1591-1598, 1993  
A:Title: Potato sucrose transporter expression in minor veins indicates a role in phloem loading  
A:Reference number: JQ2389; MUID:94146554; PMID:8312741  
A:Accession: JQ2389  
A:Molecule type: mRNA  
A:Residues: 1-516 <RTE>  
A:Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294  
A:Experimental source: cv. Desiree  
A:Comment: The gene encoding for this protein is highly expressed in mature leaves.  
C:Superfamily: common tobacco sucrose transport protein  
C:Keywords: glycoprotein; transmembrane protein  
F:31-53/Domain: transmembrane #status predicted <TM1>  
F:67-86/Domain: transmembrane #status predicted <TM2>  
F:103-122/Domain: transmembrane #status predicted <TM3>  
F:141-160/Domain: transmembrane #status predicted <TM4>  
F:180-200/Domain: transmembrane #status predicted <TM5>  
F:226-248/Domain: transmembrane #status predicted <TM6>  
F:285-304/Domain: transmembrane #status predicted <TM7>  
F:331-349/Domain: transmembrane #status predicted <TM8>  
F:366-385/Domain: transmembrane #status predicted <TM9>  
F:409-427/Domain: transmembrane #status predicted <TM10>  
F:429-448/Domain: transmembrane #status predicted <TM11>  
F:3;92/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 39.5%; Score 1079; DB 2; Length 516;  
Best Local Similarity 43.9%; Pred. No. 5.7e-78;  
Matches 225; Conservative 101; Mismatches 162; Indels 24; Gaps 11;

Qy 16 VRGAADVHVHAPISLGRILLAGVAGVQVGMALQSLTPYVOTLGLSHALTFSFMLC 75  
Db 15 VSSSLQEQPLAPSKLKIIVASIAAGVQVGMALQSLTPYVQLLGIPHKFASFWLC 74  
Qy 76 GPIAGLVQPLVGLYSDRCTSRWGRRRPFIITGCMILCVAVIVGVFSSDIGAALGDTKEH 135  
Db 75 GPISGMIVQPVGVYISDNCSSRRGRRRPFIAAGAAALVAVVFLIGFADLGHASGDT--- 131  
Qy 136 CSLYHGRWHAIIYVVLGFWLLDFSNVTQGPARAMMADLCHHGFS---AANSIFCSW 191  
Db 132 --LGKGFKPRAIAVAVVGVFWLLDVANNTLQGPCRALLADLSG--GKSGRMTANAFSFF 187  
Qy 192 MALGNILGYSGSTNNHKKWPFLLKTSACCACANLKGAFVLVAVVFLVLCITV---VTILF 248  
Db 188 MAVGNILGYAAGSYSHLFKVPFESKTRACDMYCANLKSCEFFIA-IFLLSLTTLTIALTVR 246  
Qy 249 AKEVPIRANENLPTTKAG-GEVETPTGPLAVLKGFKDLPQCMPSVLLVTAITWLSWPF 307  
Db 247 ENELPEKDEQIDEBKLAGAGSKSVFFFG--BIFGALKELPRPMWILLVLTALNWTAFWFF 304



S48789  
sucrose transport protein - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 26-May-2000  
C:Accession: S48789  
R:Buerkle, X.Y.Z.; Frommer, W.B.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S48787  
A:Accession: S48789  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-507 <BUE>  
A:Cross-references: EMBL:X82276; NID:g575350; PID:g575351  
C:Superfamily: common tobacco sucrose transport protein

Query Match 38.9%; Score 1063.5; DB 2; Length 507;  
Best Local Similarity 43.9%; Pred. No. 9.6e-157;  
Matches 225; Conservative 100; Mismatches 17; Indels 31; Gaps 13;  
Qy 19 AAVVVDHVPIS-LGRILAGWAGVQYGNALQSLTTPVOTLGLSHALTSPFMWLCGP 77  
Db 11 SSLAVEQPLPKSLKWIIMVASIAGVQFQWALQSLTTPVQLLGIDPHKFAFWLCGP 70  
Qy 78 IAGLVQPLVGLYSDRTSRNGRRPFLTGCMLICVAVVVGSSDGAALGDTKEHCS 137  
Db 71 ISGIVQPVVGYSDNCSSRFRRGFTAAAGALVTIAVFLIGFAADLGHATGD-----P 125  
Qy 138 LYHGRPHAAITVYVGLFWLLDSNTVQGPARAMMADLCHHGPSA-----ANSIFCSWMA 193  
Db 126 LKSKSPRAIAVVFVGFILVDVANNMLOGPCRALLDL---SGKARMTSNAPFSFMA 182  
Qy 194 LGNLTGSSGNTNNHWFPELKTSAACEACANLKGAFVAVFVLVCLTV---TLIFAK 250  
Db 183 VGNVLGAAGSYRLKTFPEKTPACDIYCANLKSCEFFIA-VELLSTLITALTAVVREN 241  
Qy 251 EYVRANENLPTTAKGGEVETPGPLAVLKGFDLPDGMPSVLLVTAITWLSWPPFTLY 310  
Db 242 ELPEK-DEHETDERAGARKSVPPFG-EIFGALKDLPRPMILLVLTSLNNIARFPPEFLY 299  
Qy 311 DTWMGRIYHGDPKGSNAQISAFNEGVRVGAFLGALLNSVILGFSFLIEPMCRKV-GPR 369  
Db 300 DTDMAKEVYGG--KVGDCRL--YDLGVHAGALGALLNSVVGFSMSLVEFLGKRIGGVK 355  
Qy 370 VVYTSNFMVCAAMATATISFWSLR-----DYHGVQDAITANASIKAVCLVLFPAFLGVP 425  
Db 356 RLWGLNFPVLAACVMAITVLTVMKAEKSRQYDAHGTL---MPTSGVKIGALTLEAVLGIP 412  
Qy 426 LAIYSPFVAVTAQLATRGCGGGLCTGVNLNISIVIPQVITIALGAGPMDALFGKGNIPAF 485  
Db 413 LAVTSVPFALASIFSSNAGSGGSLGLVNLAIYVPMQLVSIAGGPWDDLFGGGLNLPGE 472  
Qy 486 GVASAFALGVGVGFELLPKISKRQFRAVSAGG 518  
Db 473 IGVAVAAAAGILALTMLPSPADAKPATWGG 505

RESULT 18  
S51114  
sucrose-proton symporter - beet  
C:Species: Beta vulgaris (beet)  
C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 26-May-2000  
C:Accession: S51114  
R:Westram, A.; Eckhardt, U.; Frommer, W.B.; Riesmeyer, J.W.  
submitted to the EMBL Data Library, January 1995  
A:Description: Sequence of a sugar beet sucrose transporter cDNA.  
A:Reference number: S51114  
A:Accession: S51114  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-523 <WES>  
A:Cross-references: EMBL:X83850; NID:g633171; PID:g633172  
C:Superfamily: common tobacco sucrose transport protein

Query Match 38.2%; Score 1043.5; DB 2; Length 523;  
Best Local Similarity 43.6%; Pred. No. 3.8e-75;  
Matches 223; Conservative 84; Mismatches 160; Indels 45; Gaps 12;  
Qy 30 SLGRILLAGWAGVQYGNALQSLTTPVOTLGLSHALTSPFMWLCGPVAVLQVPLVGL 89  
Db 35 SLKKLALVASIAGVQFQWALQSLTTPVQLLGIDPHTWAPYIWLCPISGMIVQPTVGY 94  
Qy 90 YSDRCTSRWGRRRPFLTGCMLICVAVVVGSSDGAALGDTKEHCSLYHGRPHAAITV 149  
Db 95 YSDRCTSKFGRRRPFLIAGVATLVGFVSLIGFAADLGHATGDPNGV-----PKPRAITAV 149  
Qy 150 YVLGFWLIDFSNNTVQGPARAMMADLCHHGPSA-----ANSIFCSWALGNTILGYSST 205  
Db 150 FVVGFWILDVANNTLOGPCRALLDMA--AGSQAKTRYANAFSFFEMALGNIGGY-AGSY 206  
Qy 206 NNWIKWFPFLTSACCEACANLKGAFVAVVFLVCLVTLIFAKEVY-----RANENL 260  
Db 207 GRLTYVFPFTTKACDTYCANLKSCEFFISITLLVLTILALSVVRERPTLDEIQEENL 266  
Qy 261 PTTKAGGEVETPGPLAVLKG-----KDLPGMPSVLLVTAITWLSWPPFTLYDPTD 313  
Db 267 KNN-----TGGCARLPEFFQGLFALKDLKPKMLILLVLTCLNNIAPFPFLFD 316  
Qy 314 WMGRIYHGDPKGSNAQISAFNEGVRVGAFLGALLNSVILGFSFLIEPMCRKV-GPRVW 372  
Db 317 WMGKEY-----GGTVGEGKADMGVHAGALGLMINSVVLGIMSLGIEKLARLVGVKRLW 372  
Qy 373 VTSNFMVCAAMATATISFWSLRDYG--YVQDAI---TANASIKAVCLVLFPAFLGVP 427  
Db 373 GIVNLILAVCLAMTILVT-KSAEYRATHVHPGAIGPLPPGVKGGALAFVAVLGIP 431  
Qy 428 ILYSPFVAVTAQLATRGCGGGLCTGVNLNISIVIPQVITIALGAGPMDALFGKGNIPAF 487  
Db 432 ITFSIPFALASIFSSNAGSGGSLGLVNLAIYVPMQLVSIAGGPVSVTSGPMDALFGGGLNLP 491  
Qy 488 ASAFALGVGVGFELLPKISKRQFRAVSAGG 519  
Db 492 GAVATAASAILSFLLPPPPPEAKIGGSMGGH 523

RESULT 19  
A86234  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A86234  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86234  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <STO>  
A:Cross-references: GB:AF005172; NID:g2160188; PIDN:AAB60751.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: common tobacco sucrose transport protein

Query Match 36.5%; Score 996; DB 2; Length 474;  
Best Local Similarity 43.2%; Pred. No. 2e-71;  
Matches 205; Conservative 79; Mismatches 135; Indels 56; Gaps 9;  
Qy 34 LILAGWAGVQYGNALQSLTTPVOTLGLSHALTSPFMWLCGPVAVLQVPLVGLVSDR 93  
Db 11 LILAGWAGVQYGNALQSLTTPVOTLGLSHALTSPFMWLCGPVAVLQVPLVGLVSDR 93



Db 501 QVL 503

## RESULT 22

B87532

transporter, probable CC2283 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: B87532

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: B87532

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-541 &lt;STO&gt;

A;Cross-references: GB:AE005673; NID:g13423798; PIDN:AAK24254.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2283

Query Match 11.28; Score 305.5; DB 2; Length 541;

Best Local Similarity 23.98; Pred. No. 1.6e-16;

Matches 135; Conservative 76; Mismatches 211; Indels 143; Gaps 21;

Qy 14 AGVGAADVHDVAPISLGLRLLAGWAG--GVOYGWALQSLTLPYVQTGLSHALTSF 71

Db 27 AGSTGGRRGGSGMARQLSLFQIWNMCFGFGIQLGFLQANTSRFQSLGVQDVNHLAI 86

Qy 72 MWLCGPIAGLVQPLVGLYSDRCTSRMRRPFILTCMLICVAVIVVGFSSDICAALGD 131

Db 87 LWIAAPATGLLVQPLIGHFSDKTWGRGRRRPFYFWGAILTTALLVMPNSPTL----- 140

Qy 132 TKEHCSLYHGRPWAAIYVVLGFWLLDFSNNTVGGPARMMAD--LCOHHGPS--AANSIF 188

Db 141 -----WVAA-----AALIMDASINITMEPPRAFYGDNLDPDEQRATGYAMQSFF 184

Qy 189 CSMWALGNILGYSSGSTNNW--HKWFFPLKTSACCEACANLKGAFLVAVVFLVCLTVTL 246

Db 185 IG-----LGAVFASALPWLMTWNFDVANTAPAGVQPDVSRIAFYGGAGLLLAVLTV 237

Qy 247 IFAKEVPRANENLPTTKAGEV-----ETPT-----GPLAVL----- 280

Db 238 FTTRE--YSPQLTAFKAEREIAGLGLHERPEPSVNAYIALGVGVLGGAALIVWGA 295

Qy 281 -----KGFKD-----LPGMPSVLLVTAIT 300

Db 296 GLEKELYVLAGLPAFGLAGVAGARFKRIGRTDNGSFSEVLADVFRMPKTMQLAVQFFS 355

Qy 301 WLSWFPFTLYDTDMGREIYHGDPKGNAQISAFNEGVR--VGAFLGLLNLVILGFSFLLI 359

Db 356 WFLGFAMWIIYTPAVA--TVHEG---AVDASKAYNEGADWGVLFVYNGV--AALAALV 409

Qy 360 EPMCRRKVGPRVYVTSNPMFVMAATALISFWSLRYHGYVQDAITANASIKAVCLVLF 419

Db 410 IPLMVKVTSSRV-----SHAVCLGLGALGLLSFLVIRD-----PGLLWI 448

Qy 420 AFLGVPLA--ILYSVPFAVTAQLAATRGCGGGLCTGVNLISIVIPQVILALGAGPWLALF 477

Db 449 GMVGVGFAMWSILSTPYSLAGALPAR--KMGVYMGFINFVIVVPQLLAA-----TLL 499

Qy 478 GKGNIAPAGVASAFALGVGVGVFL 502

Db 500 GLMLKTFEGNQSIFALVGLGALSFAL 524

## RESULT 23

B81781

Probable integral membrane transport protein NMA2100 [imported] - Neisseria meningitidis

C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C;Accession: B81781

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre

Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A;Reference number: A81775; MUID:20222556; PMID:10761919

A;Accession: B81781

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-451 &lt;NAP&gt;

A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85316.1; PID:g738

A;Experimental source: serogroup A, strain 22491

C;Genetics:

A;Gene: NMA2100

Query Match 9.08; Score 246; DB 2; Length 451;

Best Local Similarity 22.68; Pred. No. 6.8e-12;

Matches 114; Conservative 71; Mismatches 186; Indels 134; Gaps 20;

Qy 43 GVOYGWALQSLTLPYVQTGLSHALTSFMWLCGPIAGLVQPLVGLYSDRCTSRW---- 98

Db 30 GVQTAFTLSSQMSRIFOTLGCADPHSLGWFILPLPLAGMLVQPIGVHYSR---TWKPR 86

Qy 99 -GRRRPFILTCMLICVAVIVVGFSSDICAALGDTKEHCSLYHGRPWAAIYV--VLGFWL 156

Db 87 GGRRLPYLLYGLTIAIVIMILMPSNGSGFGG-----YASL-----AALSFALMIAL 133

Qy 157 LDFSNNTVGGPARMMADLCOHHGPSAANSIFCSMWALGNILGYSSGSTNNWKKWFFPLK 216

Db 134 LDVSSNMAQPFKMWGDMVNEEQ-----GYAIGTQS-----FLA 169

Qy 217 TSACEAC-----ANL--KG-----AFLVAVVFLVCLTVTLIFAKE----- 251

Db 170 NTGAVVAAILPFVFAYIGLANTAEGVVVPQVWVAVYGAALLVITSAFTIKVKNPE 229

Qy 252 --VYPRANENLPTTKAGEVETPTGPLAVLKGFKDLPPGHPSVLLVTAITWLSWFPF-- 307

Db 230 TYARYHGIDVAANOEFKANWIELLKTAPKA-----FWTVTLVQFFCFQAFQY 275

Qy 308 -ILDYDTDMGREIYHGDPKGNAQISAFNEGVRVGAFLGL--LNSVILGFSFLEPMCR 364

Db 276 WMTYSAGAIENVWHITTDASSVGYQEGN-----WYGVLAQVSAVAIVCSFVLAKVPN 329

Qy 365 KVGPRVVMVTSNPMFVMAATALISFWSLRYHGYVQDAITANASIKAVCLVLFALGV 424

Db 330 K-----YHKAGYECCLAGALGALGPFVFFIGNQVALV-----LSVTLLGI 368

Qy 425 PLALYSVPFAVTAQLAATRGCGGGLCTGVNLISIVIPQVILALGAGPWLALFGKGNIPA 484

Db 369 AWAGIITYPLTIVTN--ALSCKHMGTYLGLFNGSICMPQIVASLLSFVLFPMLG----- 420

Qy 485 FGVASAFALGVGV-----GVFLL 503

Db 421 -GLQATMFLVGVVLLLGAFSVFLI 444

## RESULT 24

A81206

sugar transporter, probable NMB0388 [imported] - Neisseria meningitidis (strain MC58

C;Species: Neisseria meningitidis

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C;Accession: A81206

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A;Reference number: A81000; MUID:20175755; PMID:10710307

A;Accession: A81206

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-451 &lt;TET&gt;



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:29:55 ; Search time 26 Seconds  
(without alignments)  
827.931 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSGVRGAA.....VFLPKISKRFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1081	39.6	525	1	STP_SPIOL
2	424	15.5	530	1	MATP_MOUSE
3	404	14.8	530	1	MATP_HUMAN
4	135	4.9	639	1	LACY_LEULA
5	134	4.9	544	1	YD74_SYNY3
6	129	4.7	479	1	XYLP_LACPE
7	127	4.7	393	1	SOTA_ERWCH
8	126.5	4.6	618	1	SL55_RAT
9	122.5	4.5	454	1	PUCR_RHOSU
10	122.5	4.5	460	1	YAGG_ECOLI
11	122.5	4.5	721	1	YJIV_ECOLI
12	122	4.5	459	1	YDJK_ECOLI
13	120.5	4.4	460	1	VICI_ECOLI
14	120	4.4	434	1	CITL_SALTY
15	119.5	4.4	464	1	ARAE_BACSU
16	119	4.4	412	1	YICM_ECOLI
17	118.5	4.3	643	1	MOT4_CHICK
18	118.5	4.3	643	1	SL55_HUMAN
19	118	4.3	404	1	Y4XM_RHISN
20	117.5	4.3	403	1	MHPT_ECOLI
21	116.5	4.3	477	1	YPUW_RHOCA
22	116.5	4.3	522	1	STPL_ARATH
23	115	4.2	463	1	YNAJ_BACSU
24	114.5	4.2	592	1	TAT2_YEAST
25	114.5	4.2	410	1	YJIO_ECOLI
26	113.5	4.2	427	1	YDFJ_ECOLI
27	111	4.1	461	1	CSBC_BACSU
28	109.5	4.0	477	1	NOMI_PSEAE
29	109	4.0	429	1	YQ84_MYCTU
30	109	4.0	430	1	DGOT_ECOLI
31	108.5	4.0	473	1	PHDK_NOSK
32	108	4.0	469	1	MELB_ECOLI
33	107	3.9	523	1	STC_RICCO

34	106.5	3.9	393	1	TCR7_VIBAN
35	106.5	3.9	468	1	ERIC_VIBCH
36	106.5	3.9	575	1	ITR1_SCHPO
37	106.5	3.9	832	1	ATCU_SALTI
38	105	3.8	410	1	YQJV_BACSU
39	105	3.8	457	1	UIDB_ECOLI
40	105	3.8	471	1	MELB_ENTAE
41	105	3.8	587	1	LACP_KLUJA
42	104.5	3.8	391	1	YIDY_ECOLI
43	104.5	3.8	413	1	MUCK_ACICA
44	104	3.8	405	1	PMRA_LACIA
45	104	3.8	476	1	MELB_SALTY
46	104	3.8	591	1	DSDI_PSEAE
47	103.5	3.8	394	1	ARAJ_ECOLI
48	103.5	3.8	633	1	NUOL_MYCTU
49	102.5	3.8	360	1	MRAY_NEIMA
50	102.5	3.8	360	1	MRAY_NEIMB
51	102.5	3.8	394	1	EMRD_ECOLI
52	102.5	3.8	477	1	YGJI_ECOLI
53	102.5	3.8	492	1	YD19_METJA
54	102.5	3.8	615	1	UAPA_EMENI
55	102	3.7	402	1	OPDE_PSEAE
56	102	3.7	431	1	CITL_ECOLI
57	102	3.7	470	1	CYCA_ECOLI
58	102	3.7	471	1	MELB_KLEPN
59	102	3.7	522	1	STA_RICCO
60	101.5	3.7	422	1	EXUT_BACSU
61	101.5	3.7	462	1	NARU_SALTY
62	101	3.7	540	1	GTR9_HUMAN
63	101	3.7	630	1	SGA4_BOVIN
64	101	3.7	693	1	SGA3_BOVIN
65	100.5	3.7	407	1	Y661_METH
66	100.5	3.7	412	1	YWFA_BACSU
67	100.5	3.7	416	1	CHRA_PSEAE
68	100.5	3.7	546	1	Y61B_MYCPN
69	100.5	3.7	599	1	YV06_CAEEL
70	100.5	3.7	687	1	AFUB_ACTPL
71	100	3.7	389	1	YD22_DEIRA
72	100	3.7	514	1	YHK8_YEAST
73	100	3.7	620	1	SGA3_HUMAN
74	99.5	3.6	461	1	PUCR_RHOCA
75	99.5	3.6	491	1	AMPG_ECOLI
76	99	3.6	399	1	PMRA_STRPN
77	99	3.6	463	1	GUTA_BACSU
78	99	3.6	487	1	TTDT_ECOLI
79	99	3.6	492	1	ANKH_HUMAN
80	99	3.6	514	1	QACA_STRAM
81	98.5	3.6	277	1	CYST_ECOLI
82	98.5	3.6	544	1	YC9D_SCHPO
83	98.5	3.6	627	1	YHE0_YEAST
84	98.5	3.6	637	1	NU5M_STRPU
85	98.5	3.6	832	1	ATCU_SALTY
86	98	3.6	425	1	DALT_KLEPN
87	97.5	3.6	343	1	Y841_METH
88	97.5	3.6	429	1	G6PU_HUMAN
89	97.5	3.6	457	1	AROP_ECOLI
90	97.5	3.6	529	1	Y0U1_CAEEL
91	97	3.6	436	1	MNTH_DEIRA
92	97	3.6	619	1	VALL_YEAST
93	97	3.6	702	1	CSTA_ECOLI
94	96.5	3.5	331	1	YPHD_ECOLI
95	96.5	3.5	334	1	PIT_RHIME
96	96.5	3.5	367	1	BRB2_PIG
97	96.5	3.5	400	1	BMR2_BACSU
98	96.5	3.5	762	1	AVP3_HORVU
99	96	3.5	363	1	AG22_HUMAN
100	96	3.5	396	1	SOTB_ECOS7

## ALIGNMENTS

RESULT 1

STP\_SPIOL STANDARD; PRT; 525 AA.

AC Q03411.

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Sucrose transport protein (sucrose permease) (sucrose-proton symporter).

DE Spinacia oleracea (Spinach).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Magnoliophyta; Eudicotyledons; Core eudicots; Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.

OX NCBI\_TaxID=3562;

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE=Leaf;

RX MEDLINE=93099843; PubMed=1464305;

RA Riesmeyer J.W., Willmitzer L., Frommer W.B.;

RT "Isolation and characterization of a sucrose carrier cDNA from Spinacia by functional expression in yeast."

RL EMBO J. 11:4705-4713(1992).

CC -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).

CC CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.

CC -!- PATHWAY: Sucrose metabolism.

CC -!- SUBCELLULAR LOCATION: Inner membrane.

CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X67125; CAA47604.1; -

DR PIR; S28052; S28052.

DR InterPro; IPR003662; sub-transporter.

DR Pfam; PF00083; sugar\_tr.1.

DR TIGRfams; TIGR01301; GPH\_sucrose.1.

DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.

DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.

KW Transmembrane; Transport; Sugar transport; Symport.

FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 38 58 1 (POTENTIAL).

FT TRANSMEM 72 92 2 (POTENTIAL).

FT TRANSMEM 107 127 3 (POTENTIAL).

FT TRANSMEM 145 165 4 (POTENTIAL).

FT TRANSMEM 184 204 5 (POTENTIAL).

FT TRANSMEM 230 250 6 (POTENTIAL).

FT TRANSMEM 295 315 7 (POTENTIAL).

FT TRANSMEM 338 358 8 (POTENTIAL).

FT TRANSMEM 373 393 9 (POTENTIAL).

FT TRANSMEM 422 442 10 (POTENTIAL).

FT TRANSMEM 455 475 11 (POTENTIAL).

FT TRANSMEM 488 508 12 (POTENTIAL).

FT DOMAIN 509 525 CYTOPLASMIC (POTENTIAL).

SQ SEQUENCE 525 AA; 54992 MW; 0183747AD2CCLCC6 CRC64;

Query Match 39.6%; Score 1081; DB 1; Length 525;

Best Local Similarity 44.9%; Pred. No. 1.9e-70;

Matches 227; Conservative 89; Mismatches 166; Indels 24; Gaps 10;

QY 27 APISGLRLIAGWAGVQYQWALQSLTTPYVQPLGLSHALTSPMWLCGGPTAGLVQPL 86

DB 31 AEATLKLGLVASVAGVQWALQSLTTPYVQPLGLSHALTSPMWLCGGPTAGLVQPL 90

QY 87 VGLYSDRCTSRGRRRPFITGCMILVAVIVGFSSDGAALGDTKEHCSLYHGPRWHA 146

DB 91 VGYSDRCTSRGRRRPFITGCMILVAVIVGFSSDGAALGDTKEHCSLYHGPRWHA 145

QY 147 AIVYVGLFWLLDFSNVTGGPARAMMADLC--DHGSPSAANSIFCSWALGNILGYSYSGS 204

Db 146 IAVFVGVFWILDVANNTLQGPCRALLADMAAGSQTKRYANAFSFFMALGNIGYAAGS 205

QY 205 TNNHKKWFPFLKTSACCACANLKGAFVAVVFLVLCITVTLIFAKE-----VYRANEN 259

Db 206 YSRLYTVFFFTKACDVCYCANLKSCFFISITLIVLILALSVVVKERQITIDEIQEED 265

QY 260 LPT-TKAGGEVETEPTGPLAVLKGFKDLPMPSPVLLTATITLWSWFFFIYLDMDMGRE 318

Db 266 LKNNSSGSCARLPFFGQL--IGALKDLPKPKMLILLVLTALNWLAWPFLFLDMDMGKE 323

QY 319 IYHGDPKGSNAQISAFNEGVVRGAGLLNSVILGFSFLIEPCMKV-GRPVVWVTSNF 377

Db 324 VY---GGTGVGKLYDQGVHAGLGLMINSVLGVMSLSIEGLARMVGGAKRLMGIVNI 379

QY 378 MVCYAMAATALI--SFWSLRDYGHVQDAI--TANASTKAVCLVLFAPLGVPLALYSVP 433

Db 380 ILAVCLAMTLVLTSAEHRDHHIMSGAVPPPPAGVKGALAIFAVLGIPLAITFSIP 439

QY 434 FAVTAQLAATRGQGLCTGVINISIVIPQVITIALGAGPMDALFGKGNIPAFGVSAPAL 493

Db 440 FALASIFSASSSGGLSLGLVNLAIIVPQMFVSVTSGPDMFGGGLNLPFVVGVAAT 499

QY 494 VGVVGVPELLPKISKRPRAVSAGGH 519

Db 500 ASAVLSETLPLSPPPPEAKIGGSMGGH 525

RESULT 2

MATP\_MOUSE STANDARD; PRT; 530 AA.

ID MATP\_MOUSE AC P58355;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma antigen AIM1) (Underwhite protein).

GN MATP OR AIM1 OR UW.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=gdg; TISSUE=Eye, Kidney, and Uterus;

RX MEDLINE=21372467; PubMed=11479596;

RA Fukumachi S., Shimada A., Shima A.;

RT "Mutations in the gene encoding B, a novel transporter protein, reduce melanin content in medaka."

RL Nat. Genet. 28:381-385(2001).

RN [2]

RP SEQUENCE FROM N.A., AND VARIANTS UW-DBR ASN-153 AND PRO-435.

RX MEDLINE=21473748; PubMed=11574907;

RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T., King R.A., Brilliant M.H.;

RT "Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4."

RL Am. J. Hum. Genet. 69:981-988(2001).

CC -!- FUNCTION: Melanocyte differentiation antigen. May transport substances required for melanin biosynthesis (By similarity). (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By similarity).

CC -!- TISSUE SPECIFICITY: Melanocytes, eyes, kidney and uterus.

CC -!- DISEASE: Defects in MATP are the cause of the UW-dbr phenotype that results in loss of nearly all pigmentation in the homozygous state.

CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial



entitles requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF360357; AK81713.1; MGD; MGI:2153040; Matp. Disease mutation: Albinism. KW Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision; FT TRANSNEM 1 45 CYTOPLASMIC (POTENTIAL). FT TRANSNEM 46 66 1 (POTENTIAL). FT TRANSNEM 67 68 EXTRACELLULAR (POTENTIAL). FT TRANSNEM 69 89 2 (POTENTIAL). FT TRANSNEM 90 105 CYTOPLASMIC (POTENTIAL). FT TRANSNEM 106 126 3 (POTENTIAL). FT TRANSNEM 127 138 EXTRACELLULAR (POTENTIAL). FT TRANSNEM 139 159 4 (POTENTIAL). FT TRANSNEM 160 184 CYTOPLASMIC (POTENTIAL). FT TRANSNEM 185 205 5 (POTENTIAL). FT TRANSNEM 206 216 EXTRACELLULAR (POTENTIAL). FT TRANSNEM 217 237 6 (POTENTIAL). FT TRANSNEM 238 318 CYTOPLASMIC (POTENTIAL). FT TRANSNEM 319 339 7 (POTENTIAL). FT TRANSNEM 340 366 EXTRACELLULAR (POTENTIAL). FT TRANSNEM 367 387 8 (POTENTIAL). FT TRANSNEM 388 398 CYTOPLASMIC (POTENTIAL). FT TRANSNEM 399 419 9 (POTENTIAL). FT TRANSNEM 420 425 EXTRACELLULAR (POTENTIAL). FT TRANSNEM 426 446 10 (POTENTIAL). FT TRANSNEM 447 477 CYTOPLASMIC (POTENTIAL). FT TRANSNEM 478 498 11 (POTENTIAL). FT TRANSNEM 499 504 EXTRACELLULAR (POTENTIAL). FT TRANSNEM 505 525 12 (POTENTIAL). FT TRANSNEM 526 530 CYTOPLASMIC (POTENTIAL). FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL). FT VARIANT 153 153 D -> N (IN UW-DBR). FT VARIANT 435 435 S -> P (IN UW-DBR). FT SEQUENCE 530 AA; 57961 MW; F4EAD07916D9FC CRC64; Query Match 15.5%; Score 424; DB 1; Length 530; Best Local Similarity 25.3%; Pred. No. 2.6e-23; Matches 135; Conservative 89; Mismatches 196; Indels 114; Gaps 17;

QY 30 SLGRLLIAGVAGVQVGMALQISLLTPYVQTGLSHALTSFWMGLGPIAGLVVQPLVL 89  
 DB 31 STGRLLVHMSAMGREGCYAVEAYVTPVLLSVGLSPKSLYSWMLLSPIGLLPQVVS 90

QY 90 YSDRCTSRGRRRPFILTCGMLICVAVIVVGFSSDICAAL---GDTREHCSLYHGPR--- 143  
 DB 91 ASDHCRARWRRRPIILT-----LAIMML-----LGMALYNGDAVV-SALVANPROKL 138

QY 144 -WHAAYVYVGLFWLLDFSNNTVQGPARAMADLCHDRHGPSAANSIFCSWMLGNILGYS 202  
 DB 139 IWAISITMW-GVVLDFDSADFIDGPIKAYLFDVCSHQDKKGLHYHALFTGFGGALGYIL 197

QY 203 GSTNNHKKWPFPLKTSACCACANLKA-----FLVAVFLVCLTTLIFAKEVPYRAN 257  
 DB 198 GAIDMWHL-----DLGRLLGTFRQVMFFSALVLIILCFITHLCSIPAPLRDA 245

QY 258 ENLPTRK-----AGGEVETE-----PTGP-----LA 278  
 DB 246 ATDPPSQDQGSLSASGMEYGSIEKVNKGADTEQVQEWKKNKPSQSORTSMKMS 305

QY 279 VLKGFKLPPGMPSVLLVLTATLWSLPPFFLYLDWMGREIYHGDPRK--SNAQISAFNE 336  
 DB 306 LLRALVNMPSHYRLCVSHLIGMTAFSLNMLFFTDENGQIVYHGDGPGYGAHNSTEFILYER 365

QY 337 GVRVGAAGLLNSVILGESSFLIEPCKRGKVPVWVTSNFMVCMVAAATLISFWSLRD 396  
 DB 366 GVEYCGKGLCINSVSSVYSYFQKAMYSYICLKGLY-----FM-----GYLLFGLGTG 413

QY 397 YHGVQDAITANASIKAVCLVLPFAFLGVPPLAIFYVFPVTAQL-----A 441  
 DB 414 FIGLFPNYSY-----LVLCSMFGMSSTLYTPFNLTAIEYHREEKEKGOEAPGGP 465

QY 442 ATRGGGGLCTGVNLISIVIPVIAALGAGPMDALFGKNIPAFGVASAFALVG 495  
 DB 466 DNQGRGKGVDCALTCWVQLAQILVGGGLFLNMAGSVVVVVI-TASAVSLIG 518

RESULT 3  
 MATP\_HUMAN STANDARD; PRT; 530 AA.  
 AC OSUMX9; Q9BTM3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma antigen AIM1).  
 GN MATP OR AIM1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM AIM-1A).  
 RC TISSUE=Melanoma;  
 RX MEDLINE=21115844; PubMed=1121837;  
 RA Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.; "Use of an in vitro immunoselected tumor line to identify shared melanoma antigens recognized by HLA-A\*0201-restricted T cells."; Cancer Res. 61:1089-1094(2001).  
 RL [2]  
 RN SEQUENCE FROM N.A. (ISOFORM AIM-1B).  
 RC TISSUE=Skin;  
 RA Strausberg R.; Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN ALTERNATIVE SPLICING.  
 RP Ferro S.; Unpublished observations (NOV-2001).  
 RL [4]  
 RN DISEASE, AND VARIANT LEU-374.  
 RX MEDLINE=21473748; PubMed=11574907;  
 RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davissson M.T., King R.A., Brilliant M.H.; "Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4."; Am. J. Hum. Genet. 69:981-988(2001).  
 CC -!- FUNCTION: Melanocyte differentiation antigen. May transport substances required for melanin biosynthesis (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; AIM-1a (shown here), AIM-1b and AIM-1c; are produced by alternative splicing.  
 CC -!- TISSUE SPECIFICITY: Expressed in most melanoma cell lines and melanocytes.  
 CC -!- DISEASE: Defects in MATP are the cause of oculocutaneous albinism type 4 (OCA4). OCA4 is characterized by hypopigmentation of skin, hair and eyes. It leads to reduced visual acuity.  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to a frameshift in position 188.  
 CC -!- CAUTION: The described alternatively spliced isoforms are inferred using information from ests.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF172849; AAD51812.1; EMBL: BC003597; AA003597.1; ALT\_FRAME. MIM: 606202; -.  
 KW Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;

KW Polymorphism; Albinism; Alternative splicing.  
FT DOMAIN 1 46  
FT TRANSSEM 47 67  
FT DOMAIN 68 68  
FT TRANSSEM 69 89  
FT DOMAIN 90 110  
FT TRANSSEM 111 131  
FT DOMAIN 132 138  
FT TRANSSEM 139 159  
FT DOMAIN 160 184  
FT TRANSSEM 185 205  
FT DOMAIN 206 216  
FT TRANSSEM 217 237  
FT DOMAIN 238 318  
FT TRANSSEM 319 339  
FT DOMAIN 340 366  
FT TRANSSEM 367 387  
FT DOMAIN 388 398  
FT TRANSSEM 399 419  
FT DOMAIN 420 425  
FT TRANSSEM 426 446  
FT DOMAIN 447 477  
FT TRANSSEM 478 498  
FT DOMAIN 499 504  
FT TRANSSEM 505 525  
FT DOMAIN 526 530  
FT CARBOHYD 356 356  
FT VARSPLIC 129 187  
FT VARSPLIC 188 295  
FT VARSPLIC 386 406  
FT VARSPLIC 407 530  
FT VARIANT 374 374  
SQ SEQUENCE 530 AA; 58301 MW; F14A4BACA8AF31B CRC64;  
/FTID=VAR\_012162.  
Query Match 14.8%; Score 404; DB 1; Length 530;  
Best Local Similarity 24.2%; Pred. No. 7.1e-22;  
Matches 128; Conservative 88; Mismatches 20; Indels 112; Gaps 16;  
QY 33 RILAGMAGGVQYQWALQSLTTPVOTLGLSHALTSPMGLCGPIAGLVQPLVGLYSD 92  
DB 34 RLIMISMAMFGREFYAVEAAVTPLVLSVGLPSSLYSIVMPLSLPILGLQPVVGSASD 93  
QY 93 RCTSRWRRRRPPLT-GCMILCAVIVVGVGSSDGAALGDTKEHCSLYHGPR----WHA 147  
DB 94 HCRSRWRRRRPPLTGLVMVGLMALNGATVVAALIAN-----PRKLVW-AI 142  
QY 148 IVYVLGFWLLDSNTVQVOPARAMADLCHDHGSPSAANSIFCSWALNGLIYSSGSTNN 207  
DB 143 SVTMIGVVLDFDAADFIDGPIKAYLDFVCSHQDKKGLHYHALFTGFGGALYLLGAIDW 202  
QY 208 WHKWPFLKTSACCEACANLKA-----FLVAVFLVLCFLVTLFAKEV----- 252  
DB 203 AHL-----ELGRLLTEFQVFMFFTSALVLTCTFVHLCISEAPLTVAKGIP 250  
QY 253 PYRANENLPTTKAG---GEVETEPTG---PLAVLKGEK----- 284  
DB 251 PQOTQDPLSDGMYEYSIEKVNGVYNPELANOGAKNHNHAEQTRAMTLKSLRAL 310  
QY 285 -DLPGMPVSLVLTATWLSWPPFLYDMDWGREIYHGDPKG---SNAQISAFNEGVRVG 341  
DB 311 VNMPPHYRLCLISHLGTAFSLNLFETDFMGQIVYRGDPYSAHNSTEFLIYERGVVG 370  
QY 342 AGLLLNSVILGFSFLIEPCMKRGVPRVWVTSNFMVCMVAMATALI-SFWSLRDYHGY 400  
DB 371 CMGFCSINSVSSLYSFQKVLVSYTGLKGLYFTGLGFTGLGFPNYS----- 423  
QY 401 VQDAITANASIKAVCLVFAFGLVPLAILYSVPFAVTAQL-----AATRGG----- 446  
DB 424 -----TLVCLSLFGVMSSILYTPFNLIYEHREEKEKQAPGDPNSV 469  
QY 447 -GGGLCTGVNLNITSIVPOVIALGAGPNDALFGKGNIPAFGVASAFALVG 495

Db 470 RGKGMDCATLTCMVQLAQILVGGGLFLVNTAGTGVVVVVI-TASAVALIG 518  
RESULT 4  
LACY\_LEULA  
ID LACY\_LEULA STANDARD; PRT; 639 AA.  
AC Q48624;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Lactose permease (Lactose-proton symport) (Lactose transport protein).  
GN LACS.  
OS Leuconostoc lactis.  
OC Plasmid pNZ63.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=1246;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NZ6009;  
RX MEDLINE=96209221; PubMed=8633855;  
RA Vaughan E.E., David S., de Vos W.M.;  
RT "The lactose transporter in Leuconostoc lactis is a new member of the  
Lacs subfamily of galactoside-pentose-hexuronide translocators.";  
RL Appl. Environ. Microbiol. 62:1574-1582(1996).  
CC -!- FUNCTION: RESPONSIBLE FOR TRANSPORT OF BETA-GALACTOSIDES INTO THE  
CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM),  
AND ALSO FOR TRANSPORT OF HOMOLOGOUS AND HETEROLOGOUS EXCHANGE OF  
BETA-GALACTOSIDES.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- DOMAIN: THE ENZYME IIA-LIKE REGION MAY SERVE A REGULATORY  
FUNCTION.  
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
SODIUM:GALACTOSIDE SYMPORTER FAMILY (SGF).  
CC -!- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed, usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U47655; AAC44113.1; -  
DR HSP; P45618; 2GPR.  
DR InterPro; IPR001927; Na/Gal symp.  
DR InterPro; IPR001127; PTS\_EIIA.  
DR Pfam; PF00338; PTS\_EIIA\_1; 1.  
DR ProDom; PD002243; PTS\_EIIA; 1.  
DR TIGRFAMS; TIGR00792; gph; 1.  
DR TIGRFAMS; TIGR00830; PTBA; 1.  
DR PROSITE; PS00371; PTS\_EIIA\_1; 1.  
DR PROSITE; PS00872; NA\_GALACTOSIDE\_SYMP; 1.  
KW Transmembrane; Sugar transport; Symport; Phosphorylation;  
Plasmid.  
FT DOMAIN 1 473 SGF DOMAIN.  
FT DOMAIN 474 639 EIIA DOMAIN.  
FT MOD\_RES 557 557 PHOSPHORYLATION (BY HPR) (BY SIMILARITY).  
FT TRANSSEM 20 40 POTENTIAL.  
FT TRANSSEM 59 79 POTENTIAL.  
FT TRANSSEM 99 119 POTENTIAL.  
FT TRANSSEM 124 144 POTENTIAL.  
FT TRANSSEM 176 196 POTENTIAL.  
FT TRANSSEM 207 227 POTENTIAL.  
FT TRANSSEM 264 284 POTENTIAL.  
FT TRANSSEM 294 314 POTENTIAL.  
FT TRANSSEM 323 343 POTENTIAL.  
FT TRANSSEM 347 367 POTENTIAL.  
FT TRANSSEM 398 418 POTENTIAL.  
FT TRANSSEM 433 453 POTENTIAL.  
SQ SEQUENCE 639 AA; 70153 MW; 25DF2819761B415 CRC64;

Query Match 4.9% Score 135; DB 1; Length 639;  
 Best Local Similarity 22.4% Pred. No. 0.017;  
 Matches 104; Conservative 69; Mismatches 161; Indels 130; Gaps 28;

Qy 62 LGLSHALTSFNLGCPAGLVVQPLVGLYSDRCTSRWGRRRPFLITGCMCLICVAVIVVGF 121  
 Db 59 IGLTALVVIIRL-----AEVIDPILGNIVDNTKRWGKPKQVQVIGAVSVLLVVI-F 113  
 Qy 122 SSDGAALGDKTKECHSLYHGRWHA-AIVYVGLWLLD-----FSNNTVQGPARAMMADLC 176  
 Db 114 TGIFGLA-----HIN-----WIAFAIVFVFLVLLDIFYSFADVAVWGVPAISED-- 159  
 Qy 177 DHGPSAANSIFCSMMALGNILGSSGTTNNWH-----KWFPELKTSAACEACAN 226  
 Db 160 -----SKERGIFTS-----LGFTGSI-CWNGLTMIVVPVTVYFTFIATGKEHQGSG 206  
 Qy 227 LKGAFLVAVFLVCLTVLIFAEKVEPYRANENLPTTKAGGEVETPTGPLAVLKGFKDL 286  
 Db 207 WFGFSIVVSIYAVLSALAVAFGTKE-----KDLIRNAA-----TKKT-----SIKDV 249  
 Qy 287 PGMPSVLLVTAITWLSWEPFLVDTDMW---GREIYH-----GDPKGSNAQISAFNEGV 338  
 Db 250 FSG-----IHNQDLWIS-LAYLMSLAVVTVNGVLFFYEFKVLGKP-----NEFW 295  
 Qy 339 RVGAFGLLNSVLGFSFLIEPCMKRYGPRVWVTSNFMVCMVAMAATALISFWSLRDYH 398  
 Db 296 IAGAI-----ATVIGFSTAPYVPLNK-----FITRKVLFSIGOMAMIL-----SYL 337  
 Qy 399 GYVQDAITANASIKAVCLVLFAGVPLAIVLYSPFAVTAQLAATRGGGQGLCTGVNLIS 458  
 Db 338 FFIFG--KTNNMMVTIGILFNFTFAQLVVLVSLTDSIEY-----GQ-LKNGNRNEA 386  
 Qy 459 IIVQPVIIALGAGPW-DALFCKGNIPAF--GVASAFALVGGVWG 499  
 Db 387 VV-----LAVRMLDKITG-----AFSNGLVGAIAITAGMTG 418

RESULT 5  
 YD74\_SUNY3  
 ID YD74\_SUNY3 STANDARD; PRT; 544 AA.  
 AC P74158;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical symporter sll1374.  
 GN SLL1374.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=11148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hikosawa M., Suglura M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuoka A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY  
 CC (SGF).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL; D90912; BAA18257.1; -;  
 DR InterPro; IPR001927; Na/Gal\_sym.  
 DR TIGRams; TIGR00792; gph\_1  
 DR PROSITE; PS00872; NA\_GALACTOSIDE\_SYMP; FALSE\_NEG.  
 KW Hypothetical protein; Transport; Transmembrane; Symport;  
 FT Complete proteome.  
 FT TRANSMEM 31 51 POTENTIAL.  
 FT TRANSMEM 52 72 POTENTIAL.  
 FT TRANSMEM 84 104 POTENTIAL.  
 FT TRANSMEM 116 136 POTENTIAL.  
 FT TRANSMEM 162 182 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 230 250 POTENTIAL.  
 FT TRANSMEM 257 277 POTENTIAL.  
 FT TRANSMEM 318 338 POTENTIAL.  
 FT TRANSMEM 356 376 POTENTIAL.  
 FT TRANSMEM 383 403 POTENTIAL.  
 FT TRANSMEM 407 427 POTENTIAL.  
 FT TRANSMEM 450 470 POTENTIAL.  
 FT TRANSMEM 501 521 POTENTIAL.  
 SQ SEQUENCE 544 AA; 60064 MW; C91D0EDFF32277EE CRC64;

Query Match 4.9% Score 134; DB 1; Length 544;  
 Best Local Similarity 19.3% Pred. No. 0.017;  
 Matches 109; Conservative 77; Mismatches 211; Indels 168; Gaps 25;

Qy 42 GGVOYGWALQLSLTPVY-----QTLGLSHALTSEFWMLCGPIAGLVVQPLVGLYSDRCTS 96  
 Db 19 GAGDFGPAITANILVYLLFFLTDVAGVPAALAGSLVMIGKIFDINPDIIGLSDRTRS 78  
 Qy 97 RWGRRRPFILTG----CMLICVAVIVVGVSSDGAALGDTKEHCSLYHGRWHAIVVVL 152  
 Db 79 RWGRRLPWMLGMPFALFYTAQWLIHPHSD-----RLTN-QWGLFIYVVA 124  
 Qy 153 GFWLLDFSNTVQGPARAMMADLCHDHPG-SAANSIFCSWALGNILG-----YSSGST 205  
 Db 125 IMAFNLCTYTTVNLPTALTPELTQNYNERTRLNSFRFAFSIGSILSLIYLITAGLP 184  
 Qy 206 NNHKKWF-----PFLKTSACCEACANL--KG-----AFLVAVVFL 238  
 Db 185 DRPQOFGELGVMISVLSISALLWSALRQEKREPIPSLRRLAPLMAAGITLILL 244  
 Qy 239 VLCLTVTLIFAKEVPY-----RANENLPTTKAGGEVETEP--- 273  
 Db 245 AIAKSFNLGGSGEDYISFFLILLGLWGGFTLRDSAVEEHLQ-----KLENSPSPG 298  
 Qy 274 -TGPLAVLKGFKDLPPGMPSVLLVTAITWLSWFP-----FILDYDWMGREIYHGDPK 325  
 Db 299 VTENLPLKQLK-IAFSNRAFLFVIGIYLCWLAVALQTLFASILVYFVFSWM----- 347  
 Qy 326 GSNQAISAFNEGVRVGAFGLLNSVLGFSFLIEPCMKRYGPRVWVTSNFMVCMVMAA 385  
 Db 348 GLNEQOS-----GTIALAVQGTAL-VMLFVWQALAQFLDKKVIYFLGSM---VWMA 395  
 Qy 386 TALISFNSLRDHYGVQDAITANASIKAVCLVLFAGVPLAIVLYSPFAVTAQL----- 440  
 Db 396 EA--GLWLVQP--GVQ-----ALLVTLAIFAGVGVSVAYLIPWSMPDVPVLDLE 440  
 Qy 441 --AATRGG-----OGLCTGVNLISIVI-----PQVIALGAGPWDAAL 476  
 Db 441 LNTCKRRGGFFYAFMVLVQLKQVGLALGLFLVGLTLEASGFIARIPGEPIPIQPDLSALMAIR 500  
 Qy 477 FGKGNIPAFGVASAFALVGGVGVGF 501  
 Db 501 FAVAPLPAP-----FLIGGLILAIIF 520

RESULT 6  
 XYLP\_LACPE  
 ID XYLP\_LACPE STANDARD; PRT; 479 AA.  
 AC P96792;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

15-DEC-1998 (Rel. 37, Last annotation update)  
 Putative xylose-proton symporter (Xylose transporter).  
 GN XYP.  
 OS Lactobacillus pentosus.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=1589;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD353;  
 RF MEDLINE=98233724; PubMed=9573180;  
 RA Challou S., Lokman B.C., Leer R.J., Posthuma C., Postma P.W.,  
 RA Pouwels P.H.;  
 RA "Cloning, sequence analysis, and characterization of the genes  
 RT involved in isoprimeverose metabolism in Lactobacillus pentosus."  
 RL J. Bacteriol. 180:2312-2320(1998).  
 RN [2]  
 RP REVISIONS.  
 RA Challou S., Lokman B.C., Leer R.J., Posthuma C., Postma P.W.,  
 RA Pouwels P.H.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: RESPONSIBLE FOR TRANSPORT OF XULOSE INTO THE CELL, WITH  
 CC THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY  
 CC (SGF).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U89276; AAC62250.1; -;  
 DR InterPro; IPR001927; Na/Gal\_symp.  
 DR InterPro; IPR003662; sub\_transporter.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 DR TIGRFAMs; TIGR00792; gph; 1.  
 DR PROSITE; PS00872; NA.GALACTOSIDE\_SYMP; 1.  
 KW Transport; Sugar transport; Transmembrane; Symport.  
 FT TRANSMEM 54 74  
 FT TRANSMEM 102 122 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 174 194 POTENTIAL.  
 FT TRANSMEM 205 225 POTENTIAL.  
 FT TRANSMEM 253 273 POTENTIAL.  
 FT TRANSMEM 289 309 POTENTIAL.  
 FT TRANSMEM 321 341 POTENTIAL.  
 FT TRANSMEM 348 368 POTENTIAL.  
 FT TRANSMEM 397 417 POTENTIAL.  
 FT TRANSMEM 431 451 POTENTIAL.  
 SQ SEQUENCE 479 AA; 52505 MW; 1555AD1084D459C7 CRC64;  
 Query Match 4.7%; Score 129; DB 1; Length 479;  
 Best Local Similarity 18.4%; Pred. No. 0.034;  
 Matches 93; Conservative 91; Mismatches 153; Indels 168; Gaps 25;  
 YQ 50 LQLSLTPYV-----QTGLSHALTSPFWLCPAGLVVQPLVGLYSDRCTSRWGRPPF 104  
 DB 44 LSLVSLVSTYLMFFYTDVFGISAAIVGTFLVARIYDAFDGPFWMGIMDHTTRWGKSRPY 103  
 YQ 105 ILTGCM-----LICVAV-----IVVGFSSDIDGALGDKHCXLYHGRPHAAIVY 150  
 DB 104 WLWFAIPFAVSVLCFTVPNMSTGMKVVAIVTYIGVDV-----LY-----SAVNI 149  
 YQ 151 VLGFLLDFSNNTVQGPARAMMADLCDDHGPAAANSIFCSWMALGNILYSGSGSTNNWHK 210  
 DB 150 PITSILPSLTSNPQE---RVLTSLTRQPMGTIGATII--STIALPLVAYFGGSGTSSAHG 204  
 YQ 211 WPFPLKTSACCACANLKGAFVAVVFLVCLTV--TLIFAKEVYPRANENLPTTKAGCEV 269

Db 205 W-----FMVALIMAVIAMVIFVFA-----NTKERVOT-----V 234  
 QY 270 ETEPTGLAVLKGFKDLPPGMPSPVLLVTAITWLSWFPILYDTDMGRE-----IYHGD 323  
 Db 235 QSKSIFIKT--SLKALKRNPWVIVI-----FINFIY--WLGQTRSQTVYFFK 281  
 QY 324 PKGSAQISAFNEGVVRGAFGLLLNSVLGFSFLIEPMCKRGVPRVWVTSNFMVCVAM 383  
 Db 282 YNMHDATLASFILGLQLVA---LLAVVITPWT-----KRICKR-----NTMLMGMLL 326  
 QY 384 AATALLISFWSLRDYGIVQDDAITANASKACVLCVLAFLGVPPLAILYSVPPAVTAQLAAT 443  
 Db 327 AIVGQLILW-----GGSKA-----LNVPTTIVGTIVGYL----- 355  
 QY 444 RGGGGLCTGYLNISIVIPQVITIALGAGPMDALFGKGNIPAFGVASAF-----LV 494  
 Db 356 ---GTGFVSGLLIAYMLA-----DSVDYGEW-----KNCVRAEGIVTSFSSFAKFGMIG 402  
 QY 495 GGVVGVFLLPKISKRPRAVSAGGH 519  
 Db 403 GAVTGLIL-----SAGGY 415  
 RESULT 7  
 SOTA\_ERWCH STANDARD; PRT; 393 AA.  
 ID SOTA\_ERWCH AC Q9S3K0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sugar efflux transporter A.  
 GN SOTA.  
 OS Erwinia chrysanthemi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 OX NCBI\_TaxID=556;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=A350;  
 RX MEDLINE=20138160; PubMed=10671456;  
 RA Condemine G.;  
 RT "Characterization of Sota and SotB, two Erwinia chrysanthemi proteins  
 RT which modify isopropyl-beta-D-thiogalactopyranoside and lactose  
 RT induction of the Escherichia coli lac promoter."  
 RL J. Bacteriol. 182:1340-1345(2000).  
 CC -!- FUNCTION: Involved in the efflux of sugars. The physiological role  
 CC may be the reduction of the intracellular concentration of toxic  
 CC sugars or sugar metabolites. Transports IPTG, lactose and  
 CC arabinose.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Probable).  
 CC -!- INDUCTION: Activated by cyclic AMP receptor protein (CRP).  
 CC -!- SIMILARITY: BELONGS TO THE SET FAMILY OF TRANSPORTERS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ249180; CAB53452.1; -;  
 DR InterPro; IPR004750; Sugar\_eff.  
 DR TIGRFAMs; TIGR00899; 2A0120; 1.  
 KW Transport; Sugar transport; Transmembrane; Inner membrane.  
 FT TRANSMEM 22 42 POTENTIAL.  
 FT TRANSMEM 51 71 POTENTIAL.  
 FT TRANSMEM 82 102 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 152 172 POTENTIAL.  
 FT TRANSMEM 174 194 POTENTIAL.  
 FT TRANSMEM 219 239 POTENTIAL.

[illegible]

```

QY 309 LYDTDMGREIYHGDPKGSAFNEGVRVGAFLNLSVILGFSFSLIEPDMCKVGP 368
Db 348 -----YSGTLSTASTSINAM-RAVTV-----EDLIKPRMPGLAP 380
QY 369 R-VVVWTSNF-----MVCVMAAATALI-----SFWSLRDYHGYVQDATTANASIKAV 414
Db 381 RKLVEISKGLSPIYGSACLTVAALSLLGGVSGFTVMGVISGFLGGAFTLGLMLPA- 439
QY 415 CLVLAFAFGLVPLAILYSPFAVTAOL-----AATRCGGGGCLCTGVNLINISIVIPQVILALGA 470
Db 440 CNTPGVLSGLAAGLAVSUWVAGATLYPPGEOTMGVLPSTAGCTNDSVLL----- 490
QY 471 GPWDALFCKGNIPAFGV 487
Db 491 GPPGATNASNGIPSSGM 507

RESULT 9
PUCCLRHOSU
ID PUCCLRHOSU STANDARD; PRT; 454 AA.
AC P95656;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Protein pucc.
GN Pucc.
OS Rhodovulum sulfidophilum (Rhodobacter sulfidophilus).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodovulum.
OX NCBI_TaxID=35806;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97276905; PubMed=9130598;
RA Hagemann G.E., Katsiou E., Forkl H., Steindorf A.C., Tadros M.H.;
RT "Gene cloning and regulation of gene expression of the puc operon
from Rhodovulum sulfidophilum."
RL Biochim. Biophys. Acta 1351:341-358(1997).
CC -!- FUNCTION: Pucc is required for high-level transcription of the
PUC operon (by similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U81968; AAB59008.1;
DR InterPro; IPR004896; Pucc.
DR Pfam; PF03209; Pucc; 1.
KW Antenna complex; Transmembrane.
FT TRANSMEM 35 55
FT TRANSMEM 70 90
FT TRANSMEM 110 130
FT TRANSMEM 140 160
FT TRANSMEM 180 200
FT TRANSMEM 210 231
FT TRANSMEM 270 290
FT TRANSMEM 299 319
FT TRANSMEM 330 350
FT TRANSMEM 359 379
FT TRANSMEM 391 411
FT TRANSMEM 454 AA; 48432 MW; 0C2AB4ED9118C3D4 CRC64;
SQ SEQUENCE 454 AA; 48432 MW; 0C2AB4ED9118C3D4 CRC64;

Query Match 4.5%; Score 122.5; DB 1; Length 454;
Best Local Similarity 22.1%; Pred. NO. 0.094;
Matches 110; Conservative 64; Mismatches 184; Indels 139; Gaps 25;

QY 29 ISLGRLLAGVAGGVQVQGWALQLSLLTPYVOTLGLSHALTSFWMWLGPIAGLVV----- 83

```

```

Db 30 LPLSRLLRSMFQ--VSVGMAMVLLVGT-----LNRVMIVELEVPAISVIGIMISLPL 80
QY 84 ---OPLVGLYSDRCTSRWG-RRRPFILTCGML-----ICVAVIVY-GFSSDITGAALG 130
Db 81 FAPFRALIGFSDTHKTSALGRRRVPIWKGTLLQWGGFAIMPFADIVLSGQESAAQA--- 137
QY 131 DTKECHSLYHGPRHAAIYVVLGFWLLDFSNNTVQGPAPAMMADICDHGSPSAANSIFCS 190
Db 138 -----PEWIGILSAAVSFLVCAGVHTVQTVGLATATDLAPREDQPNVVGMLMV 186
QY 191 WMAIGNILGYSSGSTNNHKKWPFPLKTSACCEACANLGAFLVAVVFLVCLTVTLIFAK 250
Db 187 MLLVGMIV--SALLFGMWLEDFYHAKLIK-----IQCAAVATMVFNVIAL-----WKM 233
QY 251 EV--PYRANENLPTTKAGGEVETETPGPLAVLKGPKDLPPGMPVLLVTAITWLSWEPFI 308
Db 234 EARDVRARQRL-----EGDPEPS-----FRE-----AWGLP- 260
QY 309 LYDTDMGREIYHGDPKGSAFNEGVRVGAFLNLSVILGFSFSLIEPDMCKVGP 368
Db 261 -----TRGPNARRLLWVIGLTLGFG-SDVLLPEFGGQVLD 296
QY 369 RVWVTSNFMVCMVMAATALISF-WSLRDY-HGYVQDAIT---ANASIKAVCLVLFALG 423
Db 297 MSVAATTK--LTAAVAGGTGTVGFAMASRVLSRGVDPMMAGMAGVAVGLPAFAITFS--- 351
QY 424 VPLATLYSVFPAVTAQLAATRGGG---OGLCTGVNLINISIVIPQVILALGAGFWDALFGK 480
Db 352 ---ATLQSEPVFVGTMMAGFGAGLFSHGTLTATWRS---PKAQVGLALGAN-----G 399
QY 481 NIPAFGVASAFALVGVV 497
Db 400 AVQATSAGATIAL-GGV 415

RESULT 10
YAGG_ECOLI
ID YAGG_ECOLI STANDARD; PRT; 460 AA.
AC P75683; P71292;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical symporter yagG.
GN YAGG OR B0270.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kallman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC (SGF). STRONG, TO E.COLI YICJ.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



```

QY 362 MCRKVGPRVWVTSNFMVCMAMATATLISFWSLRDYGHVODAI-----TANAS 410
Db 535 -----TDSLVAIGITAGCVLWGLYLYQGV-DPLGGVSLWPLFGISNQM 580

QY 411 IKAVCLVL 418
Db 581 LAVALVL 588

RESULT 12
ID YDK_ECOLI STANDARD; PRT; 459 AA.
AC P76230; P76911;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical metabolite transport protein ydjk.
GN YDK OR B1775.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishiochi K., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 Genome
RT Corresponding to the 28.0-40.1 min Region on the Linkage Map.";
RL DNA Res. 3:363-377(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000272; AAC74845.1; -
DR EMBL; D90821; BAAL5573.1; ALT_INIT.
DR Ecogene; EGI3487; ydjk.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 26 46 1 (POTENTIAL).
FT DOMAIN 47 60 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 61 81 2 (POTENTIAL).
FT DOMAIN 82 90 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 91 111 3 (POTENTIAL).

```

```

FT DOMAIN 112 112 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 113 133 4 (POTENTIAL).
FT DOMAIN 134 153 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 154 174 5 (POTENTIAL).
FT DOMAIN 175 181 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 182 202 6 (POTENTIAL).
FT DOMAIN 203 271 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 272 292 7 (POTENTIAL).
FT DOMAIN 293 301 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 302 322 8 (POTENTIAL).
FT DOMAIN 323 329 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 330 350 9 (POTENTIAL).
FT DOMAIN 351 351 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 352 372 10 (POTENTIAL).
FT DOMAIN 373 399 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 400 420 11 (POTENTIAL).
FT TRANSMEM 421 441 12 (POTENTIAL).
FT DOMAIN 442 459 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 459 AA; 49602 MW; BCAB53ECB8BDD77 CRC64;

Query Match 4.5%; Score 122; DB 1; Length 459;
Best Local Similarity 22.4%; Pred. No. 0.1; Indels 140; Gaps 25;
Matches 104; Conservative 54; Mismatches 167;

QY 132 TKEHCS--LYHGP--RHAAIYVVLGFWLDFNNVTQGPARAMMADL--CDHHGPSAA 185
Db 5 TKPHCARLDRLPDCRWHSSMFAIVAFGLLVCSNAVGLLILAOALKALGWDNSTATFS 64

QY 186 SIFCSWML----GNILCYSSGSSNNHKKWPFPLKTSACCACANLKGAFLVAVVFLVC 241
Db 65 AITTAGMFLGALVGGIIGDKTGRN-----AFILYEA-IHTASVVGVGAFSPNMDFLIAC 117

QY 242 -----LTVTLIFAKEVYPYRANENLPTTKAGGEVETPTGPLA--VLKGFKDLPPG 289
Db 118 RFVKGVLGALLVYL-FAGFTEYMPGRNRGTWSSRSVFIGNWSYPLCSLIAMGLTPLISA 176

QY 290 -----MPSV--LLVTATITLSWFFPILYLDMDMGREIYHGDPKGNSNAQISAFNEG 338
Db 177 EWNVRVQLLIPAILLSLIATALAW-RYFP---ESPRWLES---RGRYQAEKVMRSIEGV 229

QY 339 --RVG-----AFGLLNSVLGFSFSLIEPCKRVGPRVV-- 371
Db 230 IROTGKPLPPVVIADGKRAPOAVPYSALLTGLVLLKRVILG-----SCVLIAANNVVOY 281

QY 372 ---WVTSNFM-----VCVAMAA-----TALISFWSLRDY 397
Db 282 TLINLWPIFTWQGINLKSIVLNTMSMFGAPFGIFIAMLVMDKIPKRTMGVGLLILAV 341

QY 398 HGYVODAITANASIKAV--CLVLFAPLGVPLAILLYSVP--FAVTAQLAATRGGGGLCTG 453
Db 342 LGYIYSLQTSMLLITLIGFLITFYMYVVCYASAVYVPEIWPTEAKLR-----GSGLANA 396

QY 454 VLNTS-IVIPQVIITAGAGPDALFGKGNIPAFGVASAFALVGGV 497
Db 397 VGRISGAAPYAVAVL-----LSSYGVTVGVFILLGAV 428

RESULT 13
YICJ_ECOLI STANDARD; PRT; 460 AA.
AC P31435; P76724;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical symporter yicJ.
GN YICJ OR B3657.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;

```



```

MEDLINE=93315143; PubMed=7686882;
*Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RNA sequence and analysis of 136 kilobases of the Escherichia coli
genome: organizational symmetry around the origin of replication.";
Genomics 16:551-561(1993).
-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(potential).
-1- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
(SGF). STRONG, TO E.COLI YAGG.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
-----
CC EMBL; L10328; AAC62010.1; ALT_INIT.
CC EMBL; AE000443; AAC76681.1; ALT_INIT.
CC EcoGene; EGI1686; y1cJ.
CC InterPro; IPR001927; Na/Gal_symp.
CC TIGRfams; TIGR00792; gph_1.
CC PROSITE; PS00872; NA_GALACTOSIDE_SYMP; 1.
CC Hypothetical protein; transport; transmembrane; Inner membrane;
CC Symport; Complete proteome.
CC FT TRANSMEM 12 32 POTENTIAL.
CC FT TRANSMEM 33 53 POTENTIAL.
CC FT TRANSMEM 81 101 POTENTIAL.
CC FT TRANSMEM 117 137 POTENTIAL.
CC FT TRANSMEM 153 173 POTENTIAL.
CC FT TRANSMEM 182 202 POTENTIAL.
CC FT TRANSMEM 249 269 POTENTIAL.
CC FT TRANSMEM 289 309 POTENTIAL.
CC FT TRANSMEM 311 331 POTENTIAL.
CC FT TRANSMEM 367 387 POTENTIAL.
CC FT TRANSMEM 404 424 POTENTIAL.
CC SEQUENCE 460 AA; 50948 MW; BBDDC8BC7D2D2B13 CRC64;
-----
Query Match 4.4%; Score 120.5; DB 1; Length 460;
Best Local Similarity 18.9%; Pred. No 0.13;
Matches 94; Conservative 66; Mismatches 187; Indels 151; Gaps 18;
-----
QY 50 LQLSLTPYVQTGLSHALTSFMWLCGPIAGLVQPLVGLYSDRCTSRGRRRPFILTCG 109
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28 VLMYMFYTDIEGIPAGVGTWFLVARALDAISDCWGLLADRTSRGKFERPWLFGA 87
QY 110 MLICVAVIVGFSDDGALGDTKEHCSLYHGPRWHAIVYVIGFWLLDFSNNTVQGP 169
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 LPFGICVCLAYSTPDL-----MNGMKIYAITYTLLTLY----- 123
QY 170 AMADLDRHGPSAANSIFCSWMAIGNILYSSGSTNNWHKFPFLKTSACCEACANLKG 229
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 -----TVNTPYC-----ALGVITNDPTQRIQLSQW----- 150
QY 230 AFLVAVVFLCLTVTLIFAKEVPYRANENLPTTKAGGEVETEPTGFLAVLKGFKDLPPG 289
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 REVLT-----AGGMLSTVLMMPLVNLIGDKNPLG 181
QY 290 MPSVLLVTAITWLSWFFPFLYDIDWNG-----REYHGDPKGSNAQISAFN- 335
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 FOGGIAVLSVVAFMMLAFCEFTTKERVEAPPTTMSREDLRDIWQNDQWRIVGLLITFNI 241
QY 336 --EGVRVGA-----FGLLLNSVILG--FSSFLIEPMCR-KVGPRVWVWTSN 376
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
242 LAVCVRGGAAMMYVYTWLGTPEVFAFLTYCYGNLIGLSALAKPLTDWKCKVITFWMTN- 300
QY 377 FMVCVMAATALISFWSLRDYGHVQDAITANASIKAVCLVLAFLAVPLAIVSVFVAV 436
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 -----ALLAVTISLAMFFVPMQASIT-----MEVFIEV-ICVLHQLVTP 338
QY 437 T-AQLAATRGGGGLCTG--VLNISIVIPQVITIALGAGPMDALFG-----KGNIPAFGVAS 489

```

CC -----BAA14228.1; -  
DR EMBL; D90203; BAA14228.1; -  
DR EMBL; AE008727; AAL19633.1; -  
DR EMBL; AL627267; CAD05152.1; -  
DR PIR; JQ0576; JQ0576.  
DR StyGene; SG10058; cita.  
DR InterPro; IPR004736; Cit\_H\_symport.  
DR Pfam; PF00083; sugar tr; 1.  
DR TIGRfams; TIGR00883; 2A0106; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 2.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Citrate utilization; Transport; Transmembrane; Inner membrane;  
KW Symport; Complete proteome.  
FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 22 42 1 (POTENTIAL).  
FT DOMAIN 43 54 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 55 75 2 (POTENTIAL).  
FT DOMAIN 76 87 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 88 108 3 (POTENTIAL).  
FT DOMAIN 109 111 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 112 132 4 (POTENTIAL).  
FT DOMAIN 133 164 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 165 185 5 (POTENTIAL).  
FT DOMAIN 186 186 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 187 207 6 (POTENTIAL).  
FT DOMAIN 208 238 7 (POTENTIAL).  
FT TRANSMEM 239 259 8 (POTENTIAL).  
FT DOMAIN 260 276 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 277 297 8 (POTENTIAL).  
FT DOMAIN 298 304 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 305 325 9 (POTENTIAL).  
FT DOMAIN 326 335 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 336 356 10 (POTENTIAL).  
FT DOMAIN 357 366 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 367 387 11 (POTENTIAL).  
FT DOMAIN 388 400 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 401 421 12 (POTENTIAL).  
FT DOMAIN 422 434 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 434 444 12 (POTENTIAL).  
SQ SEQUENCE 434 AA; 47188 MW; E0F077EAC70DB444 CRC64;  
  
Query Match 4.4%; Score 120; DB 1; Length 434;  
Best Local Similarity 20.2%; Pred. No. 0.14;  
Matches 98; Conservative 73; Mismatches 165; Indels 148; Gaps 24;  
  
QY 61 TLGLSHALTSMWLCGPIAGLVQPLVGLYSDRCTSRGRRRPFLT----GCMILICVAV 116  
DB 54 SLMLTFVFGSFLMRPVGAIV---LGAYIDRI---GRKGLMVTILAMCGCTLLIA- 104  
QY 117 IVVGFSSDIGAALGDKHECHSLYHGPRWHAIVYVGLFWLLDFSNNTVQGPAMMADLC 176  
DB 105 LVPGYQT-IGLA-----APALVLLGRLLQGSAGVGLGVSVVLSSEI- 145  
QY 177 DHHGPSAANSIFCSW-----MALGNILGYSGST-----NNWHKWFPLKTSACCE 222  
DB 146 ---ATPGKNGFYTSQMSAQVAVVAALIGYSLNITLGHDAISEW-GW----- 190  
QY 223 ACANLKGAFVAVVFLVLCVTLVIFAKEVYPRANENLPTTKAGEVEEPTGP---LAV 279  
DB 191 -----RIPFFICGMIIPLFV-----LRSQTEAEFLQKRPDTRIEFATI 233  
QY 280 LKGFKDLPPGMPSVLLVTAITWLSWFFLYDMDMGREIYHGPDKGSAOISAFNEGVR 339  
DB 234 AKNWRRIITAG--TLVAMTV--TFYFITYVTYGRTV-----LN 270  
QY 340 VCAFGLLNSVLGFSSEFLIEPMCKVKVPRVWVTSNFWCVMAATALISFWSLRYHG 399  
DB 271 LSARDSLVTLVGVSNFVILPPIGGAISDRI-----GRRVLVMTLLALITFWPMQW-- 324  
QY 400 VQVDAITANASIKAVCLVL--FAFL-----GVPLAIL-----YSVPFAVTAQLAATR 444  
DB 325 -----LTAAPDFTRMTLVLLWFSFFGNGYNGAMVAALTEVMPVIVRTVGFSLAFSLATAI 379

QY 445 GGGGGLCTGVNLISIVIPQVITIALGAGPWFALFGKINIPAFGVASAFALVGGVGVFLFP 504  
DB 380 FGG-----LTPAISTAL-----VKLTGDKSSPGWLMCA-ALCGLAATAMLEFV 421  
QY 505 KISK 508  
DB 422 RLSR 425  
  
RESULT 15  
ARAE\_BACSU STANDARD: PRT; 464 AA.  
ID ARAE\_BACSU  
AC P96710;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Arabinose-proton symporter (Arabinose transporter).  
GN ARAE.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the gram-positive bacterium Bacillus subtilis";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE OF 1-223 FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=97197523; PubMed=9045819;  
RA Sa-Nogueira S., Mota L.J.;  
RT "Negative regulation of L-arabinose metabolism in Bacillus subtilis: characterization of the arar (araC) gene";  
RN [3]  
RP SEQUENCE OF 223-464 FROM N.A., AND FUNCTION.  
RC STRAIN=168;  
RX MEDLINE=98062200; PubMed=9401028;  
RA Sa-Nogueira I., Ramos S.S.;  
RT "Cloning, functional analysis, and transcriptional regulation of the Bacillus subtilis arae gene involved in L-arabinose utilization";  
RN [4]  
RP J. Bacteriol. 179:7705-7711(1997).  
CC -1- FUNCTION: UPTAKE OF ARABINOSE ACROSS THE BOUNDARY MEMBRANE WITH

CC THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein..  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z99121; CAB15401.1; -  
 DR EMBL: X98354; CAA66998.1; -  
 DR EMBL: Y12105; CAA72812.1; -  
 DR Subtilist: BG11907; araE.  
 DR InterPro: IPR003663; CHO\_transport..  
 DR InterPro: IPR003662; sub\_transporter..  
 DR Pfam: PF00083; sugar\_tr; 1.  
 DR PRINTS: PR00171; SUGRTNSPORT..  
 DR TIGRFAMS: TIGR00879; SP; 1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 2.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
 KW Transport; Sugar transport; Transmembrane; Symport; Complete proteome.  
 FT TRANSMEM 21 43 POTENTIAL.  
 FT TRANSMEM 63 85 POTENTIAL.  
 FT TRANSMEM 92 111 POTENTIAL.  
 FT TRANSMEM 116 138 POTENTIAL.  
 FT TRANSMEM 150 172 POTENTIAL.  
 FT TRANSMEM 185 207 POTENTIAL.  
 FT TRANSMEM 266 288 POTENTIAL.  
 FT TRANSMEM 303 325 POTENTIAL.  
 FT TRANSMEM 332 354 POTENTIAL.  
 FT TRANSMEM 364 386 POTENTIAL.  
 FT TRANSMEM 398 420 POTENTIAL.  
 FT TRANSMEM 424 446 POTENTIAL.  
 SQ SEQUENCE 464 AA; 50411 MW; 13B417061CB61DA2 CRC64;  
 Query Match 4.4%; Score 119.5; DB 1; Length 464;  
 Best Local Similarity 21.3%; Pred. No. 0.16;  
 Matches 111; Conservative 67; Mismatches 174; Indels 169; Gaps 27;  
 QY 30 SLGRLLIAGVA--GGVOYQW-----ALQLSLTPYVQTLGLSHALTSFWMWLG 76  
 DB 19 SMGFVILISCAAGGLGLLYDYTAIVSAIGFLKDLVSLSPFMEGLVISSIM----- 70  
 QY 77 PIAGLVQPLVGLYSDRCTRWGRRRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHC 136  
 DB 71 -IGVGVGISGFLSD-----RFGRRK-ILMTAALLFAISAIVSALSQDVSTLI----- 117  
 QY 137 SLVHGRPRWHAIVVVLGFWLLDFSNNTVOGPARAMADLGDHGPSSAANSIFGSMWALGN 196  
 DB 118 -----IARIIGGLGI-----GMGSLSVTYITEAPPAIRGSLSLYQLFT 158  
 QY 197 ILGYSS-----GSTNNW-----HKWFPFLKTSACCEACANLGAFLVAVVFLVLC 241  
 DB 159 ILGISATYFINLAVQSGTVEGWVHTGWRWN-----LAYGMVPSVIFELVL 204  
 QY 242 LVTYTLIAKEVPY-----RANENLP-TTKAGG-----VETEPTGPLAVLK 281  
 DB 205 LVV-----PSPRLAKRAGKTNEALKTLTRINGETVAKEELKNIENTSKIEQMGSLSQL- 258  
 QY 282 GFKDLPQMPVLLVATITVLSWFPFFLYDTDMGRIYHGDPKGSNAQISAFNEGRVVG 341  
 DB 259 -FK---PGLRKALVIGILLALFNQVIGMNAITYYGPIFKMFGQNA---GFVTTICVG 311  
 QY 342 AFGLLNSVILGFSFLIEPMCRKVRVVMVTSNFM--VCVAMAATALISFWSLRDYHG 400  
 DB 312 VVEVIEFVIAV---LLIDKVRK---KLMSIGSAFAIFMILIGT---SFY----- 353  
 QY 401 VQDAITANASIKAVCLVFLAFLGVPVLAFLYVSPVAVTAQLAARGGGGGLCTGVNLNLSIV 460  
 DB 354 ----FELTSGMTIMIVLI-----LGFVAACVSV-----GPIITWIM-ISEI 388

QY 461 IPOVVIAGAG-----PWDALFGKGN-----IPAFGVASAF 491  
 DB 389 FPNHLRARAGIATIFLWGANWAIGQFVPMWIDSGLAYTF 429  
 RESULT 16  
 YICM\_ECOLI  
 ID YICM\_ECOLI STANDARD; PRT; 412 AA.  
 AC P31438; P76725;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yicm.  
 GN YICM OR B3662.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=K12 / MGI655;  
 RX MEDLINE=93315143; PubMed=7686882;  
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;  
 RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli  
 RL Genomics 16:551-561(1993).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L10328; AAA62014.1; ALT\_INIT.  
 DR EMBL: AE000444; AAC76685.1; ALT\_INIT.  
 DR EcoGene; EG11689; yicm.  
 DR InterPro: IPR003662; sub\_transporter..  
 DR Pfam: PF00083; sugar\_tr; 1.  
 KW Hypothetical protein; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 FT TRANSMEM 38 58 POTENTIAL.  
 FT TRANSMEM 71 91 POTENTIAL.  
 FT TRANSMEM 102 122 POTENTIAL.  
 FT TRANSMEM 124 144 POTENTIAL.  
 FT TRANSMEM 164 184 POTENTIAL.  
 FT TRANSMEM 192 212 POTENTIAL.  
 FT TRANSMEM 232 252 POTENTIAL.  
 FT TRANSMEM 272 292 POTENTIAL.  
 FT TRANSMEM 298 318 POTENTIAL.  
 FT TRANSMEM 322 342 POTENTIAL.  
 FT TRANSMEM 360 380 POTENTIAL.  
 FT TRANSMEM 383 403 POTENTIAL.  
 SQ SEQUENCE 412 AA; 43626 MW; 280CFD30C6912B2 CRC64;  
 Query Match 4.4%; Score 119; DB 1; Length 412;  
 Best Local Similarity 20.1%; Pred. No. 0.15;  
 Matches 107; Conservative 76; Mismatches 155; Indels 194; Gaps 28;  
 QY 6 GGQALAEISAGVGRGAADV--DHVAPISLGRILLAGMAGGVQVQWALQSLSLITVQTLG 63  
 DB 14 GNTMSEIAENRGADAITRPNWSAVSFVAFVACLII---VEF---LPVSLTTPMAODLG 67  
 QY 64 LSHALTSFWMWLCGPIAG--LVVQPLVGLYSD---RCTSRWGRRRRPFILTCMLICVAVIV 118  
 DB 68 ISEG-----VAGQSVTVTAFAVAMFASLFTQTQTQATDRYVIVLFAVLTLSCLL 117  
 QY 119 VGFSSD-----IGAALGDTKEHCSLYHGRPWHAIVVVLGFWLLDFSNNTVOGPA 168  
 DB 111

Db 118 VSFANSFSLLLGRACLGALG-----GFWMSASLTMRLVPP 155  
Qy 169 RAMMADLCDDHGPSAANSIFCS-----WMALGNILYSGSTNNHKKWFFFLKTSACC 221  
Db 156 RTV-----PKALSVIFGAVSTALVIAAPLGSFLGELIG-----W----- 189  
Qy 222 ECANKLGAFLVAVVFLVCLVTLIFAKEVYPRANENLPTTKAGGEVETETPTGPLAVLK 281  
Db 190 -----RNVFNAAAVGVLCI-----EWIITKSLP-----SLP-----GEPHOKONTFRLLQ 230  
Qy 282 GFKDLPFGMPVSLVLTATWLSWFFILYDWMGREIYHGDPKGSKNAQISAFNEGVAVG 341  
Db 231 -----RPGVMAGMIAIFMSFAGQFAFFTYI-----RPVY-----MNLAGF-----G 266  
Qy 342 AFGLLNSVLGFSFLIEPMCKRKYGPVWVTSNFM-----VGVMAATALISFWSLRDY 397  
Db 267 VDLGLVLLSFGIASF-----IGTSL-----SFLKRSVKLALACAPLI----- 306  
Qy 398 HGVYODAITANASIKAVCLVLF-----FLGVPLAILYSVPEA-----VTAQLA--AT 443  
Db 307 -----LAVSALVLTWGSVKIVATGVALLIWLGLTFALVPVGVSTWITRSLADQAE 355  
Qy 444 RGGGGLCTGVNLISIVPQVIALGAGPMDALFGKGNIPAGVASAFALVG 495  
Db 356 KAG-----SIQVAVIQLANTCCAA-----IGGYALDNIIGTSLPLMSG 393

RESULT 17  
MOT4\_CHICK  
ID MOT4\_CHICK STANDARD; PRT; 473 AA.  
AC P57788; Q9DGG24; Q9DEY5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Monocarboxylate transporter 4 (MCT 4).  
GN SLC16A3 OR MCT4.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yoon H., Philip N.;  
RT "Characterization of the chicken and mouse MCT4 genes: tissue  
distribution and relationship to other monocarboxylate transporters.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROTON-LINKED MONOCARBOXYLATE TRANSPORTER. CATALYZES THE  
RAPID TRANSPORT ACROSS THE PLASMA MEMBRANE OF MANY  
MONOCARBOXYLATES SUCH AS LACTATE, PYRUVATE, BRANCHED-CHAIN OXO  
ACIDS DERIVED FROM LEUCINE, VALINE AND ISOLEUCINE, AND THE KETONE  
BODIES ACETOACETATE, BETA-HYDROXYBUTYRATE AND ACETATE (BY  
SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane  
(By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SLC16 FAMILY OF TRANSPORTERS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF204396; AAF67524.1; -  
DR EMBL; AF308452; AAG25703.1; -  
DR InterPro; IPR004743; Mcarb.transprot.  
DR TIGRFAMs; TIGR00892; 2A0113.1.  
KW Transport; Symport; Transmembrane; Multigene family.  
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 18 38 POTENTIAL.  
FT DOMAIN 39 61 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 62 82 POTENTIAL.  
FT DOMAIN 83 91 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 92 112 POTENTIAL.  
FT DOMAIN 113 115 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 116 136 POTENTIAL.  
FT DOMAIN 137 149 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 150 170 POTENTIAL.  
FT DOMAIN 171 179 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 180 200 POTENTIAL.  
FT DOMAIN 201 231 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 232 252 POTENTIAL.  
FT DOMAIN 253 268 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 269 289 POTENTIAL.  
FT DOMAIN 290 297 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 298 318 POTENTIAL.  
FT DOMAIN 319 321 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 322 342 POTENTIAL.  
FT DOMAIN 343 358 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 359 379 POTENTIAL.  
FT DOMAIN 380 388 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 389 409 POTENTIAL.  
FT DOMAIN 410 473 CYTOPLASMIC (POTENTIAL).  
FT CONFLICT 103 103 A -> V (IN REF. 1; AGS25703).  
SQ SEQUENCE 473 AA; 51041 MW; 87B8F98EG75CC7E8 CRC64;

Query Match 4.3%; Score 118.5; DB 1; Length 473;  
Best Local Similarity 20.7%; Pred. No. 0.19;  
Matches 93; Conservative 64; Mismatches 173; Indels 120; Gaps 24;

Qy 18 GAAAVVDHVAIPSL-----GRLLIAG-MVAGGVQVGMALQLSLL-----TPYVOTL 62  
Db 2 GAVVDDGDGPGVKAPDGGWGNVLFQCFITIGFSYAFFKAVSVFFKELIREFGVGYSDTA 61  
Qy 63 GLSHALTSEFMWLCGPIAGLVQPLVGLYSDRCTSEWRRRPFIITGCMILCVAVIVVGF 122  
Db 62 WISSILLAMLYGTGPLCSV-----CVNRFQ-CRPVNLVGGVLFASMGMAISFC 108  
Qy 123 SDIGAALGDTKEHSCYHGRWHAIAVYVGLFWLDFSNNTVQGPARAMMADLCDDHGP- 181  
Db 109 TSI-----VQIY-----LTAGVITGLG-ALNFQ-----PSLIMLNRYFDKRRL 147  
Qy 182 -----SAANS--IFCSWMALGNILYSGSTNNHKKWFFFLKTSACCEACANLKGAFVAV 235  
Db 148 ANGLSAGSPVFLCALSPGLQILQHEYG-----W-----RGGFLILG 184  
Qy 236 VFLVCLVTLIFAKEVYPRANENLPTTKAGGEVETETPTGPLAVLKGFKDLPPGMPVLL 295  
Db 185 GMLNCCVCGALMRPLEPPKSE---ATREPAEKAKK--KLLDPSVFQD--GGFVIYTL 237  
Qy 296 VTATWLSWFFILYDWMGREIYHGDPKGSKNAQISAFNEGVAVGAFGLLNSVLGFS 355  
Db 238 AASIMVLGLFVPPVSVSY-AKDLGYQDTK-----AAFL-----TILGFI 277  
Qy 356 SFLIEPMCKRKYGPVWVTSNFMVGVMAATALISFWSLRDYHGVYQDQAITANASIKAVC 415  
Db 278 DIFARPCGMVA-GLKWKVRRP---CVYLFSAFI-----FNGFTD--LMGSMVSDYGG 324  
Qy 416 LVFLA-FUGVPLAILYSVPEAVTAQLAATR 444  
Db 325 LVVCFIFGIGYGMVGAQFVFLMAIVGTQ 354

RESULT 18  
SL55\_HUMAN STANDARD; PRT; 643 AA.  
ID SL55\_HUMAN  
AC Q92911; Q43702; Q9NWB6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Sodium/iodide cotransporter (Na(+)/I(-) cotransporter) (Sodium-iodide  
symporter) (Na+/I-symporter).  
GN SLC5A5 OR NIS.  
OS Homo sapiens (Human).



```
FT VARIANT 556 556 /FTID=VAR_010269.
FT S -> Q.
FT /FTID=VAR_010270.
SQ SEQUENCE 643 AA: 68666 MW: 02D361A27B2FDA43 CRC64;

Query Match
Best Local Similarity 4.3%; Score 118.5; DB 1; Length 643;
Matches 118; Conservative 72; Mismatches 208; Indels 179; Gaps 25;

QY 29 ISLGRLLIAGVAGG-----VOYGMALQSLTTPYVOTLGL-SHAL--- 68
DB 26 VSGTGLVGLARGQORAEFTGGRRLAALPVGLSLASFMSA-VQVLGVPSAYRG 84
QY 69 TSGFMLC-----GPIAGLVQPL---VGLXS--DRCTSRMGR-----RPFIL 106
DB 85 LKFLMWCLQLLNSVLTALLFMPFVYRLGTSTYVLEMRFSRAVRLCGTLQYIVATMLY 144
QY 107 TCGMLICAVI---VVGSSDGAALGDTKEHCSYHGPRAHAIYV-----VLGF 154
DB 145 TGIVYAPALILNQVTGL--DIWALLSTGIICTFTYAVGGMKAVVWTDVQVVMVLSGF 202
QY 155 WLDFESNTVQGPARAMMA-----DLCDHGSPSAANSIFCSNMALGNILGYSSGSTN 206
DB 203 WVVLARGVNLVGGPROVLTLAQNSHRINLMDPNPDRSRYTFWTFVVGTLWLSMYGVN 262
QY 207 -----NWHKWFPLKTSACCACANLKGAFVAVVFLVCLTV 244
DB 263 QAQORYVACRPEKQAKLALLINQVGLFVSSAACCG-----IVMFVYTDGCP 313
QY 245 TLIFAKEVPYRANENLPTTKAGGEVET-EPGCPVAVLKGFKDLPDGMPSVLLVTAITWLS 303
DB 314 LL-----GRISAPDQYMLPLVLDIFEDL-PCVPLGLFLACA----- 347
QY 304 WPFILYDIDMWGREIYHGDGPKGNAQISAFNEGVGVGAFGLLNSVILGSSFLIEPMC 363
DB 348 -----YSGTILSTASTSINAM-AAVTV-----EDLIKRL 375
QY 364 RKVGRVWVWTSNFMVCMVAMAATLISFWSLRDHYGVQDAITANASIKAVCLVLFALG 423
DB 376 RSLAPKRLVITSKGLSLVGSACLTVAALSLLGGVGLQSGFTVMGVISGPLLGAFL-ILG 434
QY 424 VPLAILYSVPFAVTAQLAATRGGGGGLCTGVNLINISIVIPQVILALGAGPWALFGKNIP 483
DB 435 MFL-----PACNTPGVLAGLGAAL-----SLWVALGATLYPPSEOTMRVL 476
QY 484 AFGVASAFAL---VGVGVGVFLLPKISKROQRAVSAG 517
DB 477 PSSAARCVALSVNASGLDLPALLP---ANDSSRAPSSG 511

RESULT 19
Y4XM_RHISN
AC P55705; STANDARD; PRT; 404 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical transport protein Y4XM.
GN Y4XM.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(Potential).
```

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative 3-hydroxyphenylpropionic acid transporter.  
 GN MHP1 OR B0353.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 ON NCBI\_TaxID=562;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=37426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
 RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Nishimoto H., Saito N.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE OF 83-330 FROM N.A.  
 RC STRAIN=K12 / CS520;  
 RA Fernandez A., Garcia J.L., Diaz E.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: COULD BE A TRANSPORTER FOR 3-PHENYLPROPIONATE  
 CC (HYDROCINNAMIC ACID).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
 CC FRAMESHIFTS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE000142; AAC73456.1; ALT INIT.  
 CC EMBL: U73857; AAB18077.1; ALT INIT.  
 CC EMBL: D85613; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: X97543; CAA66145.1; -;  
 CC Ecogene: EG13293; mhpt.  
 CC InterPro: IPR004746; Bnzte transporter.  
 CC InterPro: IPR003662; sub transporter.  
 CC Pfam: PF00083; sugar tr. 1.  
 CC TIGRFAMs: TIGR00895; 2A0115; 1.  
 CC PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
 CC PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
 CC Transport: Transmembrane; Inner membrane; Symport; Complete proteome.  
 CC DOMAIN 1 16  
 CC TRANSMEM 17 37  
 CC DOMAIN 38 53  
 CC TRANSMEM 54 74  
 CC DOMAIN 75 81  
 CC TRANSMEM 82 102  
 CC DOMAIN 103 105  
 CC TRANSMEM 106 126  
 CC DOMAIN 127 142  
 CC TRANSMEM 143 163

FT DOMAIN 164 164 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 165 185 6 (POTENTIAL).  
 FT DOMAIN 186 217 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 218 238 7 (POTENTIAL).  
 FT DOMAIN 239 253 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 254 274 8 (POTENTIAL).  
 FT DOMAIN 275 279 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 280 300 9 (POTENTIAL).  
 FT DOMAIN 301 306 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 307 327 10 (POTENTIAL).  
 FT DOMAIN 328 339 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 360 360 11 (POTENTIAL).  
 FT DOMAIN 361 369 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 370 390 12 (POTENTIAL).  
 FT DOMAIN 391 403 CYTOPLASMIC (POTENTIAL).  
 FT CONFLICT 295 295 L->V (IN REF. 4).  
 SQ SEQUENCE \* 403 AA; 41550 MW; 85EE921E57686FEB CRC64;  
 Query Match 4.3%; Score 117.5; DB 1; Length 403;  
 Best Local Similarity 19.6%; Pred. No. 0.19;  
 Matches 103; Conservative  
 Qy 37 AGYAGGV-----QYGNALQLSLTPVQTLGLSHALTSMWLCGPIAGLVVQPLV 87  
 Db 34 AGIAGGIAQAQAFALDKMQMGWIFSAGI-----LGL-----LPGALVG----- 70  
 Qy 88 GLYSDRCTSRWRRRPEI---LTGCMILICVAV-----IVVGFSSDIGAALGDT 132  
 Db 71 GMLAD-----RYRRKRLIGSVLFLSLATAIAMDFFSLVFLVRLMTGVG--LGAALPNL 124  
 Qy 133 KECHSLYHGRHAAIYVYVGLFWLLDFSNNTVQGPARAMMADLCHDHGPGSAANSTFC-- 190  
 Db 125 IALTSEAGPRFRG-----TAVSLMYCGVP 149  
 Qy 191 -WMALGNILGYSSGSTNNHKKWF-----PFLKTSACCEACANLKGALVAVVFLVCLT 243  
 Db 150 IGAALATLGF-AGANLAWQTVFWVGVVP-----LILVPLLMRWLP 190  
 Qy 244 VTLIFAREVPYRANENLPTTKRAGEVEPTGPIVLAKGFKDLPGMPFVLLVTAITWL- 302  
 Db 191 ESAVFA-----GEKQAP--PLRALFA-----PETATATLLMLWC 223  
 Qy 303 -----SWFPFITYDTDMGREIYHGDPKGSNAQISAFNEGVYRVGAFGLLLNSV 350  
 Db 224 YFTLLVYVYMLINLPLLLVBOGF-----QPSQAAGVVMFALQMGAAAG-- 266  
 Qy 351 ILGSSFLIEPCMKRVGPRVWVTSNFMVYVMAATAILISWLSRDYHYVQDATTANAS 410  
 Db 267 -----TLMGLAMDKLRP-----VTMSLLIYSGMLA-SULLALGTVSSFNGLLAGEVAG-- 314  
 Qy 411 IKAVCLVLFAFLGVPLATLYSV-PFAVTAQLAATRGSGGLCTGVLTNI--STIVPOVITA 467  
 Db 315 -----LFATGG--QSVLYALAPLFYSSQIRAT-GVCTAVAVGRIGLAMSGLAKMLA 364  
 Qy 468 LGAGPMDALFGKGNIPAFGVASAF-----LVGGVGVFLLPKIKSRKQ 510  
 Db 365 LGTG-----TVGVMAASAPGILVAGLAVFILMRSRIQ 398  
 RESULT 21  
 YPUM\_RHOCA  
 ID YPUM\_RHOCA STANDARD: PRT; 477 AA.  
 AC P26176;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 50.4 kDa protein in PUHA-BCHM intergenic region (ORF477)  
 DE (Protein F1696).  
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 ON NCBI\_TaxID=1061;  
 RN [1]





```
FT TRANSMEM 24 44 1 (POTENTIAL).
FT TRANSMEM 84 104 2 (POTENTIAL).
FT TRANSMEM 118 138 3 (POTENTIAL).
FT TRANSMEM 141 161 4 (POTENTIAL).
FT TRANSMEM 170 190 5 (POTENTIAL).
FT TRANSMEM 203 223 6 (POTENTIAL).
FT TRANSMEM 296 318 7 (POTENTIAL).
FT TRANSMEM 325 345 8 (POTENTIAL).
FT TRANSMEM 349 369 9 (POTENTIAL).
FT TRANSMEM 385 405 10 (POTENTIAL).
FT TRANSMEM 431 451 11 (POTENTIAL).
FT TRANSMEM 454 474 12 (POTENTIAL).
FT DOMAIN 475 522 CYTOPLASMIC (POTENTIAL).
FT CONFLICT 333 333 A -> G (IN REF. 1).
SQ SEQUENCE 522 AA; 57610 MW; 813AFC8AFA7AA583 CRC64;

Query Match 4.38; Score 116.5; DB 1; Length 522;
Best Local Similarity 23.8%; Pred. No. 0.29;
Matches 70; Conservative 34; Mismatches 85; Indels 105; Gaps 17;

Qy 16 VRGAAVVDHVPISLGRILLAGVAGVQVYQWA-----LQLSLTPYVQ----- 60
Db 245 IRG-----VDDVSO-EFDDLVAAKESQSIHPWNLRRKYPHLPHTWAVMIPFFQQLTGI 299

Qy 61 -----TGLSHALTSFWMCGPIAGL--VVQPLVGLYSDRCTSRWRRRPPFIL 106
Db 300 NVIMFYAPVLPNTIGFT---TDSILMSAVVTGSVNVAAATLSIYG---VDRWGRRLFLLE 353

Qy 107 TGC-WLICVAVIVVGFSSDIIAGLDGDKYKCHSLYHGRPHAAIV-----YVLGF----- 154
Db 354 GGTOMLICQAVV-----AACIGAKFGVDGTPGEL---PKWAIYVVVTFICIVVAGFAWSWG 406

Qy 155 ---WLLDPSNNTVQGPARAMADLCHRHGPSAANSIFCSWALGNILYSGSGSTNNHKKW 211
Db 407 PLGNLV-----PSEIFPLEI-----RSAQSI-----TVSVNMI 435

Qy 212 PFPLKTSACCBACANLK-GAFVAVVFLVLCVTLVTLIFAKEVPYRANENLPTTK 264
Db 436 PTFIIAQIFLTMCLHKLFGLEVFVAFVVMVSIFVIF-----LPETK 478

RESULT 23
ID YNAJ_BACSU STANDARD; PRT; 463 AA.
AC P94488;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical symporter ynaJ.
GN YNAJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Borchert S., Klein C., Piksa B., Hammelmann M., Entian K.D.;
RT "Sequencing of a 26 kb region of the Bacillus subtilis genome downstream of spoVJ."
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchsi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
```

```
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Potwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC (SGF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U66480; ABA41090.1; -
CC EMBL: 299113; CAB13641.1; -
CC Subtilist; BG12262; ynaJ.
CC InterPro: IPR001927; Na/Gal_symp.
CC TIGRFAMS; TIGR00792; gph; 1.
CC PROSITE; PS00872; NA_GALACTOSIDE_SYMP; 1.
CC Hypothetical protein; Transport; Transmembrane; Symport;
CC Complete proteome.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 334 354 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
SQ SEQUENCE 463 AA; 51534 MW; B5F0733048023FF2 CRC64;

Query Match 4.2%; Score 115; DB 1; Length 463;
Best Local Similarity 19.9%; Pred. No. 0.33;
Matches 95; Conservative 63; Mismatches 171; Indels 148; Gaps 20;

Qy 54 LLTFYVQTLGLSHALTSEFMWLCGPIAGLVQPLVGLYSDRCTSRWRRRPPILTGC-----M 110
Db 35 LLFFYTDVFGLSAAAGTFLVWRIIDALADPFITGTVIDRTNSRFARFPRLLFGAFPFV 94

Qy 111 LICVAVIVVGFSSDIIAGLDGDKYKCHSLYHGRPHAAVYVGLFWLDFSNNTVQGPARA 170
Db 95 ILAILCFTTPDFSDMGKLI-----YAYITYV-----GLSLTYTTINVPYGA 135

Qy 171 MMADLC-DHHGPSAANSIFCSWALGNIL-----GYSSGSTNN-----WHKWFPELK 216
Db 136 LTSAMTRNNQEWVSITSVRMFLANGLGLVAVFAFFVPLLAAYLSDTSGNSLQMTMGLG 195

Qy 217 TSACCACANLKGAFLVAVVFLVLC-----TVTLIFAKE-----VPRANENLPTT 263
Db 196 MIGGC-----LLIFCFKSTKERVTLQKSEEKIKETDIFEQFRVNR----- 235
```

CC 264 KAGGEVETPTGLAVLKGFKDLPGMPSVLVLTWLS-----WP----- 305  
CC 236 -----PLVLVSIFIIIFGVNSISVGIYVYTLNLERDLVKWYGLIGSLPA 283  
CC 306 -----PFILYDWMGREIYHGDPKGSKNAQISAFNEGVVRGAFGLLNSVILGFSFLIEP 361  
CC 284 LVLPFIPRLHQFLGKK-----KLLNVALLN--IIGLLALLFVP 321  
CC 362 MCRKGVPRVWVTSFMVCMVMAAATLIS-----FWSL---RDYHGYVODAITANASIKAV 414  
CC 322 -----PSNVYL---ILVRLTAAGSLTAGGYMALLPETIEYGEYRTGKRMGLIYAI 372  
CC 415 CLVLRAF---LGVPIALLSVFPVTAQLAATRGSGGQLCTGVNLNISVIVPVIALL 468  
CC 373 IGFHFKFGMALGVVPGVLVDRGVVANAQAT---PAALMGILITTTIIPVFLVLV 425

RESULT 24  
TAT2\_YEAST  
AC P38967; STANDARD: PRT; 592 AA.  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Tryptophan permease (Tryptophan amino acid transporter).  
GN TAT2 OR TAP2 OR SCM2 OR LTC3 OR SAB2 OR YOL020W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JK9-3D;  
RX MEDLINE=95021187; PubMed=7523855;  
RA Schmidt A., Hall M.N., Koller A.;  
RT "Two FK506 resistance-conferring genes in Saccharomyces cerevisiae,  
RT TAT1 and TAT2, encode amino acid permeases mediating tyrosine and  
RT tryptophan uptake.";  
RL Mol. Cell. Biol. 14:6597-6606(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94335876; PubMed=8058037;  
RA Chen X.H., Xiao Z., Fitzgerald-Hayes M.;  
RT "SCM2, a tryptophan permease in Saccharomyces cerevisiae, is  
RT important for cell growth.";  
RL Mol. Gen. Genet. 244:260-268(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
KC STRAIN=YN 140;  
RA Kawamura D., Yamashita I., Nimi O., Toh-E A.;  
RT "Cloning and nucleotide sequence of a gene conferring ability to grow  
RT at a low temperature on Saccharomyces cerevisiae tryptophan  
RT auxotroph.";  
RL J. Ferment. Bioeng. 77:1-9(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Shin Y.H., Goo D.M., So I.S., Rhode P.R., Campbell J.L., Kim J.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-481 FROM N.A.  
RA Habbig B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 156-592 FROM N.A.  
RA Hughes B., Pohl T.M.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR HIGH-AFFINITY TRYPTOPHAN TRANSPORT.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: X79150; CAA5777.1; -;  
CC EMBL: L33461; AAA60324.1; -;  
CC EMBL: U66304; BAA03811.1; -;  
CC EMBL: U66834; AAB07526.1; -;  
CC EMBL: Z74761; CAA99019.1; -;  
CC EMBL: Z74762; CAA99020.1; -;  
CC PIR: S48084; S48084.  
CC PIR: S47926; S47926.  
CC SGD: S0005380; TAT2.  
CC InterPro: IPR002293; AA/rel\_pmease1.  
CC InterPro: IPR004840; AAC\_permease.  
CC InterPro: IPR004841; Permease.  
CC InterPro: IPR004762; Yeast\_AA\_perm.  
CC Pfam: PF00324; aa\_permeases; 1.  
CC TIGRfam: TIGR00913; 2A0310; 1.  
CC PROSITE: PS00218; AMINO\_ACID\_PERMEASE\_1; 1.  
KW Transport; Amino-acid transport; Transmembrane.  
FT TRANSMEM 88 108 POTENTIAL.  
FT TRANSMEM 114 134 POTENTIAL.  
FT TRANSMEM 161 181 POTENTIAL.  
FT TRANSMEM 193 213 POTENTIAL.  
FT TRANSMEM 227 247 POTENTIAL.  
FT TRANSMEM 267 287 POTENTIAL.  
FT TRANSMEM 307 327 POTENTIAL.  
FT TRANSMEM 359 379 POTENTIAL.  
FT TRANSMEM 405 425 POTENTIAL.  
FT TRANSMEM 429 449 POTENTIAL.  
FT TRANSMEM 473 493 POTENTIAL.  
FT TRANSMEM 515 535 POTENTIAL.  
FT CONFLICT 187 187 Q -> T (IN REF. 3).  
SQ SEQUENCE 592 AA; 65404 MW; A233BAD16264B319 CRC64;

Query Match 4.2%; Score 114.5; DB 1; Length 592;  
Best Local Similarity 20.7%; Pred. No. 0.46;  
Matches 104; Conservative 60; Mismatches 138; Indels 201; Gaps 28;

QY 25 HVAPTSGLRLILAGMVAG-----GVQYGNALQSLTTPVQTLGLSHALTSEFW 73  
DB 86 HLIMTAIGSGTGTGLFVSGKRAIAGGGLGVVIGWAIAGS-----QIIGTHGL----- 134  
QY 74 LCGPIAGLVVQPLVGLYSDRCTS-----RWGRRRPFILTCMLICVAVI 117  
DB 135 --GEIT--VRFPVGAFAFYGTRELDPSISFVWSTIYVLQW-----FFVLPLEIIAAAMT 185  
QY 118 VVGFSSTGALGDTKEHCSLYHGPRWHAIVY-----VLGFWLLDFSNNTVQGP 167  
DB 186 VOYWNSSIDPVI-----WVAIFYAVIVSINLFGVRGFGAEAFSTIKAI 230  
QY 168 ARAMADLC-----DH-----HGPSAANSIFCSMMALGNILGYS----- 201  
DB 231 TVCGFIILCVVLICGGPDHPEIGAKYWHDPCLANGEPGVLSVLVASYSLGGTEMTC 290  
QY 202 --SGSTNNHKKWPFELKTSACCEACANLKGAFVAVVFLVLCITVTLIFAKEVPYRANENL 260  
DB 291 ASGETD--PKGLP-----SAIKQVFWRLFFFLISLTVLGFL--VPY-TNQL 333  
QY 261 PTTKAGGEVETPTGLAVLKGFKDLPGMPSVLVLTWLS----- 298  
DB 334 ---LGGSSVDNSPFVIAIKLHHIKALPSIVNAVILSVLSVSGNSCIFASSRTLCMSAHQ 390  
QY 299 -ITWLSWEPFLLYDFDWMGREIYHGDPKGSKNAQISAFNEGVVRGAFGLLNSVILGFSF 357  
DB 391 LIPW--WFGYI---DRAGRPL-----VGMANS-LFGLLAF 420  
QY 358 LIETPCMRKGVPRVWVTSFMVCMVMAAATLISFW-SLRDHYGYVQDAITANAS----- 410  
DB 421 LV-----KSGS--MSEVENLWMAIAGLATCIV--WLSINLSHIFRLAMKAGKSLDELE 471

```

Qy 411 -IKAV-----CLVLF 420
Db 472 FVSAVGINGSAYSALINCLILIA 494

RESULT 25
YJIO_ECOLI
ID YJIO_ECOLI STANDARD; PRT; 410 AA.
AC P39386;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transport protein yjio.
GN YJIO OR B4337.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLACASE FAMILY).
CC -1- SIMILARITY: STRONG, TO E.COLI MDFA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14003; AAA97233.1;
CC EMBL; AE000504; AAC77293.1;
CC EcoGene; EG12576; yjio.
CC Pfam; PF00083; sugar_tr.1.
CC Hypothetical protein; Transport; Transmembrane; Inner membrane;
CC Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 217 237 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 308 328 POTENTIAL.
FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
SQ SEQUENCE 410 AA; 44687 MW; 17C909971AD788B7 CRC64;

```

```

Query Match 4.2%; Score 113.5; DB 1; Length 410;
Best Local Similarity 19.9%; Pred. No. 0.38;
Matches 100; Conservative 82; Mismatches 160; Indels 161; Gaps 29;

Qy 44 VOYGWALQSLTTPVOTLGLSHALTSP-----MWLCGPITAGLVVQPLVGLYS 91
Db 19 ILYDFAAYLS--TDLIQP-GIINVVRDFNADVSLAPAAVSLYLAG---GMALQWLGLPLS 72

Qy 92 DRCISRWRGRRPFLITGCMILICVAVIVVGFSSD-----IGALGDTKEHCSLYHGPRWHA 146
Db 73 DRI-----GRRPVLITGALFTLACAAATMTTSMTQFLIARAIOGT-SICFI-----A 119

```

Search completed: April 17, 2003, 11:30:39  
Job time : 35 secs

```

Qy 147 AIVYVLGEWLLDFSNNTVQ---GPARA---MMADLCHDHGHPSPAANSIFCSWMALGNILGYS 201
Db 120 TVGYV-----TVQEAFGQTKGIKMA-----IITSIVLIAPILG 155

Qy 202 SGST-----NNWHKWPFLKTSACCEACANLKGAFVAVVFLVCLTVTLIFAKEVPYRAN 257
Db 156 SGAALMHFMHWKVLFAII-----AVMGFISFVGLLA----- 187

Qy 258 ENLPTRKAGGEVETETGTPGLAVLKGFKDL-----PPGMPSVLLVTAITWLSWFFPIL 309
Db 188 --MPETVKRGAV---PPSAKSVLRDRFNVFCNRLFLFGAAATISUSIYIPMWSWAVSPVIL 242

Qy 310 YTDWMGREIYHGDPKGSNAOISAFNEGVRCGAFG---LLNSVILGFSFSLTEPMCRKVG 367
Db 243 ID-----AGSLTTSQF-AWTQVPVFGAVIVANAIV---ARFVKDP---TE 280

Qy 368 PRVWVWTSNFMVCMAMAATALISFWSLDRYHYGVODATANASIKAVCLVLFAFLGVPLA 427
Db 281 PRFIWRA---VPIQLVGLSLLIIVGNLLSPHWLWSVLGTS-----LYAF---GIG 324

Qy 428 ILYSVPFVAVTAQLAATRCGGGGLCTGVLN-LSIVIPQVIVIIALGAGPWDALSGKGNIPAFG 486
Db 325 LIPTLFRFT---LFSNKLPRGTVSASLNMVLMVMSVSVIEGRWLW---FNGGRLP--- 375

Qy 487 VASAFALVGGVGVVFLLPKISKR 509
Db 376 -FHLLAVVAGVIVVFTLAGLLNR 397

```



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: April 17, 2003, 11:30:00 ; Search time 91 Seconds  
(without alignments)  
1175.148 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELASAGRGAA.....VFLPKISKRQFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rviris.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2552	93.4	521	10	Q9SXM0 zea mays (m
2	2325	85.1	538	10	Q9LKH3 oryza sativ
3	2317.5	84.9	537	10	Q49838 oryza sativ
4	2296.5	84.1	523	10	Q9M422 hordeum vul
5	2295.5	84.1	522	10	Q8S4W1 triticum ae
6	2287.5	83.7	523	10	Q8RUL3 triticum ae
7	2282.5	83.6	522	10	Q8S4W0 triticum ae
8	1862	68.2	506	10	Q948L0 oryza sativ
9	1862	68.2	506	10	Q944W2 oryza sativ
10	1551	56.8	612	10	Q9SQK5 vitis vinif
11	1549.5	56.7	604	10	Q9FVL6 lycopersico
12	1523	55.8	594	10	Q80605 arabidopsis
13	1312.5	48.1	297	10	Q8S7X2 oryza sativ
14	1144	41.9	533	10	Q41152 ricinus com
15	1126	41.2	501	10	Q65929 daucus caro
16	1122.5	41.1	512	10	Q9S7Z5 apium grave

Q9FE59 arabidopsis	510	10	Q9FE59
Q8VYX4 brassica ol	513	10	Q8VYX4
Q9ZTB9 apium grave	512	10	Q9ZTB9
Q9M3R4 arabidopsis	510	10	Q9M3R4
Q40938 plantago ma	510	10	Q40938
Q9XHL6 pisum sativ	524	10	Q9XHL6
Q9SP63 vitis vinif	501	10	Q9SP63
Q9SLN7 daucus caro	515	10	Q9SLN7
Q65803 daucus caro	515	10	Q65803
Q9FNR6 daucus caro	515	10	Q9FNR6
Q40777 vicia faba	523	10	Q40777
Q9FG00 arabidopsis	491	10	Q9FG00
Q9FV92 solanum tub	488	10	Q9FV92
Q9SP14 alonsoa mer	502	10	Q9SP14
Q9ZVK6 arabidopsis	492	10	Q9ZVK6
Q9Z232 arabidopsis	513	10	Q9Z232
Q9FVJ6 lycopersico	500	10	Q9FVJ6
Q9SQK6 vitis vinif	501	10	Q9SQK6
Q9C6H8 arabidopsis	491	10	Q9C6H8
Q9Z231 arabidopsis	512	10	Q9Z231
Q9F1X9 arabidopsis	492	10	Q9F1X9
Q80550 arabidopsis	512	10	Q80550
Q9M423 hordeum vul	506	10	Q9M423
Q8RW6 arabidopsis	512	10	Q8RW6
Q8VYX3 brassica ol	508	10	Q8VYX3
Q43653 solanum tub	516	10	Q43653
Q9M535 euphorbia e	530	10	Q9M535
Q40937 plantago ma	503	10	Q40937
Q9C8X2 arabidopsis	512	10	Q9C8X2
Q9SQK4 vitis vinif	505	10	Q9SQK4
Q9FNY9 arabidopsis	509	10	Q9FNY9
Q40583 nicotiana t	507	10	Q40583
Q94G12 oryza sativ	344	10	Q94G12
Q9SP15 asarina bar	510	10	Q9SP15
Q39438 beta vulgar	523	10	Q39438
Q9SM24 beta vulgar	539	10	Q9SM24
Q04516 arabidopsis	474	10	Q04516
Q9XFM1 nicotiana t	521	10	Q9XFM1
Q40167 lycopersico	429	10	Q40167
Q8W534 vitis vinif	445	10	Q8W534
Q65883 ricinus com	334	10	Q65883
Q9AVR7 ricinus com	306	10	Q9AVR7
Q8W401 nicotiana t	136	10	Q8W401
Q94JM9 oryza sativ	135	10	Q94JM9
Q94JM8 oryza sativ	262	10	Q94JM8
Q9SVC5 betula verr	599	5	Q9SVC5
Q9SVS1 drosophila	576	13	Q9SVS1
Q90Z74 oryzias lat	553	3	Q90Z74
Q14091 schizosacch	553	3	Q14091
Q9HEX4 pneumocysti	566	3	Q9HEX4
Q9A612 caulobacter	541	16	Q9A612
Q961T2 homo sapien	553	4	Q961T2
Q95K15 macaca fasc	553	6	Q95K15
Q9SEAL apium grave	157	10	Q9SEAL
Q95KC5 macaca fasc	10.3	6	Q95KC5
Q9JSW6 neisseria m	451	16	Q9JSW6
Q9K111 neisseria m	451	16	Q9K111
Q9RVB3 deinoococcus	454	16	Q9RVB3
Q8R7F4 thermoanaer	434	16	Q8R7F4
Q9U610 schizosacch	132	3	Q9U610
Q9C716 arabidopsis	77	10	Q9C716
Q9V209 pyrococcus	452	17	Q9V209
Q9SXU7 cicer ariet	78	10	Q9SXU7
Q8TZJ0 pyrococcus	442	17	Q8TZJ0
Q9C1N0 penicillium	119	3	Q9C1N0
Q9ULN3 homo sapien	618	4	Q9ULN3
Q9V210 pyrococcus	430	17	Q9V210
Q9A5Y2 caulobacter	418	16	Q9A5Y2
P70187 homo musculus	490	11	P70187
Q96NY0 homo sapien	490	4	Q96NY0
Q96MC6 homo sapien	490	4	Q96MC6
Q9DBS0 mus musculus	147.5	5.4	Q9DBS0
Q9Y2W3 homo sapien	533	4	Q9Y2W3

```

90 140.5 5.1 448 16 Q8RGR4 Q8gr4 fusobacteri
91 139 5.1 508 2 Q93KC2 Q93kc2 erwinia chr
92 138 5.1 465 10 Q9MAN5 Q9man5 arabidopsis
93 137.5 5.0 447 2 Q03095 Q03095 streptomyce
94 137.5 5.0 473 17 Q97213 Q972y3 sulfolobus
95 136.5 5.0 394 16 Q92479 Q9z479 agrobacteri
96 136.5 5.0 789 16 Q8UGX6 Q8ugx6 agrobacteri
97 136 5.0 464 11 Q9DC37 Q9dc37 mus musculu
98 136 5.0 558 3 Q3UWV5 Q3umv5 coprinus ci
99 133.5 4.9 399 16 Q8XSH5 Q8xsh5 ralstonia s
100 133.5 4.9 548 3 Q9P3K6 Q9p3k6 neurospora

```

## ALIGNMENTS

```

RESULT 1
Q9SXM0
ID Q9SXM0 PRELIMINARY; PRT; 521 AA.
AC Q9SXM0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Sucrose transporter.
GN ZMSUT1.
OS zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Aoki N., Hirose T., Takahashi S., Ono K., Ishimaru K., Ohsugi R.;
RT "Molecular cloning and expression analysis of a gene for a sucrose
RT transporter in maize (Zea mays L.).";
RL Plant Cell Physiol. 0:0-0(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AB008464; BAA83501.1; -
DR InterPro; IPR000847; HTH_LysR.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 521 AA; 55174 MW; 50E38A5983B4F881 CRC64;

Query Match 93.4%; Score 2552; DB 10; Length 521;
Best Local Similarity 92.4%; Pred. No. 2.7e-189;
Matches 484; Conservative 16; Mismatches 16; Indels 8; Gaps 3;

QY 1 MARGDGGQLAELSAGVRG-----AAAVVDHVAPISLGRILLAGMVAGGVQYGMALQSL 55
DB 1 MARGDGE--LELSVGRGTGGGAAAAAADHVAPISLGRILLAGMVAGGVQYGMALQSL 58
QY 56 TPYVQTGLSHALTSEFMILCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCMLICVA 115
DB TPYVQTGLSHALTSEFMILCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCMLICVA 118
QY 59 TPYVQTGLSHALTSEFMILCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCMLICIA 118
DB TPYVQTGLSHALTSEFMILCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCMLICIA 118
QY 116 VTIVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 175
DB VTIVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 175
QY 119 VTIVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 178
DB VTIVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 178
QY 176 CDHGHPSAANSIFCSWMALGNILYSSGSTNNHKKWFFFLKTSACCEACANLKGAFLVAV 235
DB CDHGHPSAANSIFCSWMALGNILYSSGSTNNHKKWFFFLKTSACCEACANLKGAFLVAV 238
QY 236 VFVLVCLVTLIFAKEVEPYRANENLPTTKAGEVEVETETGFLAVLKGFKDLPPGMPSVLL 295
DB VFVLVCLVTLIFAKEVEPYRANENLPTTKAGEVEVETETGFLAVLKGFKDLPPGMPSVLL 295
QY 239 VFVLVCLVTLIFAKEVEPYRANENLPTTKAGEVEVETETGFLAVLKGFKDLPPGMPSVLL 297
DB VFVLVCLVTLIFAKEVEPYRANENLPTTKAGEVEVETETGFLAVLKGFKDLPPGMPSVLL 297
QY 296 VTAITWLSWFFPIYDMDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSVILGFS 355
DB VTAITWLSWFFPIYDMDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSVILGFS 357
QY 298 VTGLTWLSWFFPIYDMDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSVILGFS 357
DB VTGLTWLSWFFPIYDMDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSVILGFS 357

```

```

QY 356 SFLEPMDCKRVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHYVODAITANASIKAVC 415
DB SFLEPMDCKRVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHYVODAITANASIKAVC 417
QY 416 LVLFAPFLGVLAILYISVFPFAVTAQLAATRGGGGGLCTGVLNISIVIPQVITIALGAGPMDA 475
DB LVLFAPFLGVLAILYISVFPFAVTAQLAATRGGGGGLCTGVLNISIVIPQVITIALGAGPMDA 477
QY 476 LFGKGNTPAGVASAFALVGGVGVFLLPKISKRQFRVAVSAGGH 519
DB LFGKGNTPAGVASAFALVGGVGVFLLPKISKRQFRVAVSAGGH 521

```

## RESULT 2

```

Q9LKH3
ID Q9LKH3 PRELIMINARY; PRT; 538 AA.
AC Q9LKH3;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Sucrose transporter.
GN SUT1.
OS Oryza sativa (indica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=39946;
RN [1]
RP SEQUENCE FROM N.A.
RA Whitfield P.R., Furbank R.T.;
RT "Sucrose transporter gene from rice (Oryza sativa indica cv. IR36).";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280050; AAF90181.1; -
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 538 AA; 56157 MW; 8DC0E7712A2167A1 CRC64;

Query Match 85.1%; Score 2325; DB 10; Length 538;
Best Local Similarity 81.2%; Pred. No. 1e-171;
Matches 436; Conservative 39; Mismatches 42; Indels 20; Gaps 3;

QY 1 MARGD-----GGQLAELSAGVRG-----AAAVVDHVAPISLGRILLAGMVA 41
DB 1 MARGSGAGGGGGGGGLELSVGVGGGARGGGGGGAAAVETAAPISLGRILLAGMVA 60
QY 42 GGVOYGMALQSLTTPYVQTGLSHALTSEFMILCGPIAGLVVQPLVGLYSDRCTSRWGR 101
DB GGVOYGMALQSLTTPYVQTGLSHALTSEFMILCGPIAGLVVQPLVGLYSDRCTSRWGR 120
QY 102 RPFLTGMCLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSN 161
DB RPFLTGMCLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSN 180
QY 121 RPYLTGCVLICLVAVVIGFSADIGYAMGDTKEDCSYHGSRWHAIVVVLGFWLLDFSN 180
DB RPYLTGCVLICLVAVVIGFSADIGYAMGDTKEDCSYHGSRWHAIVVVLGFWLLDFSN 180
QY 162 NTVOGPARAMMADLCDHGHPSAANSIFCSWMALGNILYSSGSTNNHKKWFFFLKTSACC 221
DB NTVOGPARAMMADLCDHGHPSAANSIFCSWMALGNILYSSGSTNNHKKWFFFLKTSACC 240
QY 222 EACANLKGAFLVAVVFLVCLVTLIFAKEVEPYRANENLPTTKAGEVEVETETGFLAVLK 281
DB EACANLKGAFLVAVVFLVCLVTLIFAKEVEPYRANENLPTTKAGEVEVETETGFLAVLK 299
QY 282 GFKDLPMPMSVLLVTAITWLSWFFPIYDMDMGREIYHGDPKGSNAQISAFNEGVRVG 341
DB GFKDLPMPMSVLLVTAITWLSWFFPIYDMDMGREIYHGDPKGSNAQISAFNEGVRVG 359
QY 300 GFRNLPTGMPSVLLVTLGTLWLSWFFPIYDMDMGREIYHGDPKGTDPQIEAFNQGV 359
DB GFRNLPTGMPSVLLVTLGTLWLSWFFPIYDMDMGREIYHGDPKGTDPQIEAFNQGV 401
QY 342 AFGLLLSNVILGFSFLIEPMDCKRVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHY 401
DB AFGLLLSNVILGFSFLIEPMDCKRVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHY 419
QY 360 AFGLLLSNVILGFSFLIEPMDCKRVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHY 419
DB AFGLLLSNVILGFSFLIEPMDCKRVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHY 461
QY 402 QDAITANASIKAVCLVFAFLGVPLAILYSVPFAVTAQLAATRGGGGGLCTGVLNISIVI 461
DB QDAITANASIKAVCLVFAFLGVPLAILYSVPFAVTAQLAATRGGGGGLCTGVLNISIVI 461

```



Db 480 LFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 523

## RESULT 5

Q8S4W1 ID Q8S4W1 PRELIMINARY; PRT; 522 AA.  
 AC Q8S4W1;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Sucrose transporter SUT1A.  
 OS Triticum aestivum (wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,  
 RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.;  
 RT "Three sucrose transporter genes are expressed in the developing grain  
 of wheat.";  
 RL Plant Mol. Biol. 0:0-0(2002).  
 DR EMBL: AF408842; AAM13408.1; -;  
 DR EMBL: AF408844; AAM13410.1; -;  
 SQ SEQUENCE 522 AA; 55072 MW; C980802E5593A6A CRC64;  
 Query Match 84.1%; Score 2295.5; DB 10; Length 522;  
 Best Local Similarity 81.2%; Pred. No. 1.9e-169;  
 Matches 428; Conservative 41; Mismatches 45; Indels 13; Gaps 3;  
 QY 1 MARGDGGQLAELSAGVRGAAAA-----VVDHVAPISLGRLLIAGMVGAGVQYGMALQQL 52  
 DB 1 MARGGNGEVELSVGCGGGAGAGADAPVD-----ISLGRLLIAGMVGAGVQYGMALQQL 56  
 QY 53 SLITPYVOTLGLSHALTSMWLCGPIAGLVQPLVGLYSDRCTSRWGRRRRPFILTGCMLI 112  
 DB 57 SLITPYVOTLGLSHALTSMWLCGPIAGLVQPLVGLYSDRCTSRWGRRRRPFILTGCMLI 116  
 QY 113 CVAVIVVGFSSDIGAALGDSKECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMM 172  
 DB 117 CIAVVVGFSSDIGAALGDSKECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMM 176  
 QY 173 ADLCHGHPGSAANSIFCSWMAALGNILGYSSGSTNNHKKWPFPLKTSACCACANLKGAF 232  
 DB 177 ADLSAQHGPSAANSIFCSWMAALGNILGYSSGSTNNHKKWPFPLKTSACCACANLKGAF 236  
 QY 233 VAVVFLVCLTVTLIFAKEVPRANENLPTTKAGEVEETPTGFLAVLKGFKDLPPGMP 292  
 DB 237 VAVVFLVCLTVTLIFAKEVPRANENLPTTKAGEVEETPTGFLAVLKGFKDLPPGMP 295  
 QY 293 VLLVTAITWLSWFPFILDYDWMGREIYHGDPKGSNAQISAFNEGVRVGFGLLINSVIL 352  
 DB 296 VLLVTAITWLSWFPFILDYDWMGREIYHGDPKGSNAQISAFNEGVRVGFGLLINSVIL 355  
 QY 353 GFSSFLIPEMCKVKGVPRVWVTSNFMVCVMAATALISFWSLRDYHGVDQAITANASIK 412  
 DB 356 GFSSFLIPEMCKVKGVPRVWVTSNFMVCVMAATALISFWSLRDYHGVDQAITANASIK 415  
 QY 413 AVCLVLEAFGLVPLAILYVSPFAVTAQLAATRGGGGLCTGVNLISIVIPQVITAGAGP 472  
 DB 416 IVSLALFAFLGIPLAILYVSPFAVTAQLAATRGGGGLCTGVNLISIVIPQVITAGAGP 475  
 QY 473 WDALFKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 519  
 DB 476 WDELFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 522  
 RESULT 6  
 Q8RUL3 ID Q8RUL3 PRELIMINARY; PRT; 523 AA.  
 AC Q8RUL3;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Sucrose transporter SUT1D.  
 OS Triticum aestivum (wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,  
 RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.;  
 RT "Three sucrose transporter genes are expressed in the developing grain  
 of wheat.";  
 RL Plant Mol. Biol. 0:0-0(2002).  
 DR EMBL: AF408845; AAL90455.1; -;  
 DR EMBL: AF408844; AAM13410.1; -;  
 SQ SEQUENCE 523 AA; 55232 MW; AD42EBDB3B7F734F CRC64;  
 Query Match 83.7%; Score 2287; DB 10; Length 523;  
 Best Local Similarity 81.1%; Pred. No. 8.6e-169;  
 Matches 425; Conservative 42; Mismatches 51; Indels 6; Gaps 2;  
 QY 1 MARGDGGQLAELSAGVRG-----AAVVDHVAPISLGRLLIAGMVGAGVQYGMALQQL 55  
 DB 1 MARGGNGEVELSVGCGGGAGAGQPAVDISLGRLLIAGMVGAGVQYGMALQQL 60  
 QY 56 TPYVOTLGLSHALTSMWLCGPIAGLVQPLVGLYSDRCTSRWGRRRRPFILTGCMLCVA 115  
 DB 61 TPYVOTLGLSHALTSMWLCGPIAGLVQPLVGLYSDRCTSRWGRRRRPFILTGCMLCVA 120  
 QY 116 VIVVGFSSDIGAALGDSKECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 175  
 DB 121 VIVVGFSSDIGAALGDSKECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 180  
 QY 176 CDHGHGHPGSAANSIFCSWMAALGNILGYSSGSTNNHKKWPFPLKTSACCACANLKGAF 235  
 DB 181 SAQHGPSAANSIFCSWMAALGNILGYSSGSTNNHKKWPFPLKTSACCACANLKGAF 240  
 QY 236 VFLVCLTVTLIFAKEVPRANENLPTTKAGEVEETPTGFLAVLKGFKDLPPGMP 295  
 DB 241 LFLAFLVCLTVTLIFAKEVPRANENLPTTKAGEVEETPTGFLAVLKGFKDLPPGMP 299  
 QY 296 VTAITWLSWFPFILDYDWMGREIYHGDPKGSNAQISAFNEGVRVGFGLLINSVILGFS 355  
 DB 300 VTGLTWLSWFPFILDYDWMGREIYHGDPKGSNAQISAFNEGVRVGFGLLINSVILGFS 359  
 QY 356 SFLTEPMCKVKGVPRVWVTSNFMVCVMAATALISFWSLRDYHGVDQAITANASIKAVC 415  
 DB 360 SFLTEPMCKVKGVPRVWVTSNFMVCVMAATALISFWSLRDYHGVDQAITANASIKAVC 419  
 QY 416 LVLEAFGLVPLAILYVSPFAVTAQLAATRGGGGLCTGVNLISIVIPQVITAGAGPMDA 475  
 DB 420 LVLEAFGLVPLAILYVSPFAVTAQLAATRGGGGLCTGVNLISIVIPQVITAGAGPMDA 479  
 QY 476 LFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 519  
 DB 480 LFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 523  
 RESULT 7  
 Q8S4W0 ID Q8S4W0 PRELIMINARY; PRT; 522 AA.  
 AC Q8S4W0;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Sucrose transporter SUT1B.  
 OS Triticum aestivum (wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]



RP SEQUENCE FROM N.A.  
RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,  
RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.,  
RT "Three sucrose transporter genes are expressed in the developing grain  
of wheat.";  
RL Plant Mol. Biol. 0:0-0(2002).  
DR EMBL: AF408843; AAL13409.1; -.  
SQ SEQUENCE 522 AA; 55168 MW; 20EC33438765D734 CRC64;

Query Match 83.6%; Score 2282.5; DB 10; Length 522;  
Best Local Similarity 81.08; Pred. No. 1.9e-168; Indels 13; Gaps 3;  
Matches 421; Conservative 41; Mismatches 46; Indels 13; Gaps 3;

Qy 1 MARGDGGGGLAELASAGVRAA-----VVDHVAPISLGLIILAGMVAGGVQVGMALQL 52  
Db 1 MARGGNGEVELSVGGGGAGAGADAPVD----ISLGLIILAGMVAGGVQVGMALQL 56

Qy 53 SLLTPYVOTGLSHALTSFMWLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCM 112  
Db 57 SLLTPYVOTGLSHALTSFMWLCGPIAGLVVQPCVGLYSDRCTSRWGRRRPFIITGY 116

Qy 113 CVAVTVGFSDDIGAALGDTREKCHSLYHGRWHAIAIVVLGFWLLDFSNNTVOGPAR 172  
Db 117 CIAVVVGFSADIGAALGDSKECSLYHGRWHAIAIVVLGFWLLDFSNNTVOGPAR 176

Qy 173 ADLCHDHGSPSAANSIFCSWMLGNIILGYSSGSTNNHKKWFFPFLKTSACCEACANL 232  
Db 177 ADLSAQHGSPSAANSIFCSWMLGNIILGYSSGSTNNHKKWFFPFLTRACCEACANL 236

Qy 233 VAVVFLVLCVLTIFAKEVYRANENLPTTKAGGEVETETGTLAVLKGFDLPVGMPS 292  
Db 237 VAVFLAFVLCVLTIFAKEVYRANENLPTTKAGGEVETETGTLAVLKGFDLPVGMPS 295

Qy 293 VLLVTAITWLSWFFELVDTDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSV 352  
Db 296 VLLVTLWLSWFFELVDTDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSV 355

Qy 353 GFSSFLIEPMCKRKVPRVYVNTSMFVMAAATLISFWSLDRYHGYYVQDAITANAS 412  
Db 356 GFSSFLIEPLCKRLKLPVWVSSNLFVLSMAAIIISWATQDMGHYQIHAIKREIK 415

Qy 413 AVCLVLEAFVLPVLAIVSVPFATAQLAATRGGGGLCTGVLNLSIVIPVITAGL 472  
Db 416 IVSLLAFVLPVLAIVSVPFATAQLAATRGGGGLCTGVLNLSIVIPVITAGL 475

Qy 473 WDLFGKGNIPAFGVASAFALVGVVFLPKISKROFRAVSAGGH 519  
Db 476 WDLFGKGNIPAFGVASAFALVGVVFLPKISKROFRAVSAGGH 522

RESULT 8  
Q948L0 ID Q948L0 PRELIMINARY; PRT; 506 AA.  
AC Q948L0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Sucrose transporter.  
GN OSSUT3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE; TISSUE=PANICLE;  
RA Hirose T., Scofield G.N., Whitfield P.R., Aoki N., Furbank R.T.,  
RA Terao T.;  
RT "Isolation and characterisation of a cDNA for a novel sucrose  
transporter, OSSUT3, from rice.";  
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB071809; BAB68368.1; -.  
DR

DR InterPro; IPR003662; sub\_transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR TIGRFAMs; TIGR01301; GPH\_sucrose; 1.  
SQ SEQUENCE 506 AA; 52776 MW; 0844DC10E1E63E75 CRC64;

Query Match 68.2%; Score 1862; DB 10; Length 506;  
Best Local Similarity 70.3%; Pred. No. 6.1e-136;  
Matches 362; Conservative 47; Mismatches 86; Indels 20; Gaps 6;

Qy 11 ELSAGVRCGAAAVDHDVAP--ISLGLIILAGMVAGGVQVGMALQSLTPYVQVTLCLSHAL 68  
Db 6 ELDDGGDGGK-----APQISLGLFLACMVAGGVQVGMALQSLTPYVQVTLGIPHAL 60

Qy 69 TSFMMWLCGPIAGLVVQPLVGLYSORCTSRWGRRRPFIITGCMICVAVIVVGFSSD 128  
Db 61 TSFMMWLCGPIAGLVVQPCVGLYSKCTSSLGRRRRPFIITGCIICISIVIVGFSSD 120

Qy 129 LGDTKECHSLYHGRWHAIAIVVLGFWLLDFSNNTVOGPARMMADLCHDHGSPSA 188  
Db 121 LGDTTDCVYRGPRYHAAAFAILGFWLLDFSNNTVOGPARALMADLSGRHGSPSA 180

Qy 189 CSWMLGNIILGYSSGSTNNHKKWFFPFLKTSACCEACANLKGAFVAVVFLVLCVLT 248  
Db 181 CSWMLGNIILGYSSGSTNDHKKWFFPFLTRACCEACANLKAFLVAVVFLGLSTAVT 240

Qy 249 AKVEPYRANENLPTTKAGGEVETETGTLAVLKGFDLPVGMPSVLLTATITWLSW 308  
Db 241 AREVA-----LDPVAAAKRNEGEASGLLAVFKGNKLPVGMPSVLIITGLT 294

Qy 309 LYDTDMGREIYHGDPKGSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCKRK 368  
Db 295 LFDTDMGREIYHGRPDGSPAEVTAQEGVRQAGFGLLSIVILGSISSFLIEPMCKR 354

Qy 369 RVVWVTSNFMVMAAATLISFWSLDRYHGYYVQDAITANA---SIKAVCLVLEAF 425  
Db 355 RAVWVSSAVVMAAATLISFWSLDRYHGYYVQDAITANA---SIKAVCLVLEAF 414

Qy 426 LAIVSVPFATAQLAATRGGGGLCTGVLNLSIVIPVITAGLGPWDALFGKGNIPAF 485  
Db 415 FAVLCSVPFATAQLAASRGGGGLCTGVLNLSIVIPVQMAIALGAGPWDELFGEG 474

Qy 486 GVASAFALVGVVGVVFLPKISKROFRAVS-AGGH 519  
Db 475 AMASVEFAAAAAAGVLLPKVS---VRSVSMAGGH 506

RESULT 9  
Q944W2 ID Q944W2 PRELIMINARY; PRT; 506 AA.  
AC Q944W2;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Sucrose transporter.  
OS Oryza sativa (indica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=39946;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. IR 36;  
RA Whitfield P.R., Aoki N., Scofield G.N., Hirose T., Terao T.,  
RA Furbank R.T.;  
RT "Isolation and characterisation of a putative sucrose transporter gene  
(OSSUT3) from rice.";  
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF419298; AAL14982.1; -.  
DR InterPro; IPR003662; sub\_transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR TIGRFAMs; TIGR01301; GPH\_sucrose; 1.  
SQ SEQUENCE 506 AA; 52774 MW; A4E73029D7022B64 CRC64;

Query Match 68.2%; Score 1862; DB 10; Length 506;  
Best Local Similarity 69.9%; Pred. No. 6.1e-136;  
Matches 360; Conservative 48; Mismatches 87; Indels 20; Gaps 6;

```
QY 11 ELSAGVRGAAVVDHAP--ISLGRLLIAGMAGVGVQGWALQSLTTPYVQTGLSHAL 68
DB 6 ELDDGGDGKGR-----APQISLGLFLACMAGVGVQGWALQSLTTPYVQTGLSHAL 60
QY 69 TSFMLCGPIAGLVQPLVGLYSRDKTSRWGRRRPFILTGCMLCICVAVIVGVFSSDIGAA 128
DB 61 TSMWLCGPIAGLIVQPCVGLYSRDKTSRGLRRRPFILTGCIICISVIVGVFSSDIGVA 120
QY 129 LGDTKEHCSLYHGPRHRAIVVGLFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSTF 188
DB 121 LGDATEDCKVYRGPYRHAATAFILGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSTF 180
QY 189 CSWMAIGNLIGYSSGTSNNHKKWFFPKTSACCACANLKGAFVAVVFLVCLVTLIF 248
DB 181 CSWMAIGNLIGYSSGTSNDHKKWFFPKTSACCACANLKGAFVAVVFLVCLVTLIF 240
QY 249 AKFVYRANENLPTTKAGEVETETGPLAVLKGKDLPPGMPSVLLVTAITWLSWFFPI 308
DB 241 AREVA-----LDPVAAKRNKEGEASGPLAVFKGMKLPVGMPSVLLVTAITWLSWFFPI 294
QY 309 LYDTDMGREIYHGPCKGNSAQISAFNEGVYRGAFGLLINSVILGFSSFLIEPCKRKVP 368
DB 295 LFDTDMGREIYHGPCKGNSAQISAFNEGVYRGAFGLLINSVILGFSSFLIEPCKRRLGA 354
QY 369 RVVYTSNPMVCVMAATALISFWSLDYHGVQDAITANA----SIKAVCLVLFALFVLP 425
DB 355 RAVVMSSAVVCVMAAVSVLSAWSLDFGGVQDAARAPAEQGVRSALALFVFLGLP 414
QY 426 LAILSVFPAVTAQLAATRGGGGLCTGVNLISIVIPVITIALGAGPMDALFGKGNIPAF 485
DB 415 FAVLCSVPFVTAQLTASRGGGGLCTGVNLISIVIPVITIALGAGPMDALFGKGNIPAF 474
QY 486 GVSALFVGVGVVFLFPKISKROFRAYS-AGGH 519
DB 475 ANASVFAAAAAGVLLPKVS---VRSVSMAGGH 506
```

## RESULT 10

```
Q9SQK5 ID Q9SQK5 PRELIMINARY; PRT; 612 AA.
AC Q9SQK5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Putative sucrose transporter.
GN VWSUC12.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Vitales.
OX NCBI_TaxID=29760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHIRAZ; TISSUE=FRUIT;
RA Davies C., Wolf T., Robinson S.P.;
RT "Three putative sucrose transporters are differentially expressed in grapevine tissues."
RL Plant Sci. 147:93-100(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF021809; AAF08330.1; -
DR TIGRFAMs: TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 612 AA; 65711 MW; 8FC3FBD6AB439078 CRC64;
```

Query Match 56.8%; Score 1551; DB 10; Length 612;  
Best Local Similarity 54.8%; Pred. No. 8.5e-112;  
Matches 298; Conservative 78; Mismatches 110; Indels 58; Gaps 6;

QY 31 LGRLLIAGMAGVGVQGWALQSLTTPYVQTGLSHALTFSFMLCGPIAGLVQPLVGLY 90

```
DB 63 LRTILSCMIAAGVQVQGWALQSLTTPYVQTGLSHALTFSFMLCGPIAGLVQPLVGLY 122
QY 91 SDRCTSRGRRRPFLLTGCMLCICVAVIVGVFSSDIGALGDTKEHCSLYHGPRHRAATVY 150
DB 123 SDKCSRYGRRRPFLLAGSLMISVAVTIIGFSADIGYLLGDTNMDCKRFKGTRTWAAIIF 182
QY 151 VLGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSTFCSWMAIGNLIGYSSGTSNNHKK 210
DB 183 VLGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSTFCSWMAIGNLIGYSSGTSNNHKK 242
QY 211 WFPFLKTSACCACANLKGAFVAVVFLVCLVTLIFAKEVP----- 253
DB 243 WFPFLKTSACCACANLKGAFVAVVFLVCLVTLIFAKEVP----- 302
QY 254 -----YRANEN-----LPTTKAGEVETETPGPLAY 279
DB 303 NPQIGFDSNKSCLDMSAVDNATGNPNPESSEYINKNAKHLTPVIOEQNESFSDGPGAVLV 362
QY 280 --LKGFKDLPPGMPSVLLVTAITWLSWFFPIYDMDGREIYHGDGPKGNSAQISAFNEG 337
DB 363 NLLTSRLHLPAMHSVLLVMAISWLSWFFFLFDTDMGREVYHGDGPKGDSANVAYDAG 422
QY 338 VRVGAFLGLLSVILGFSSFLIEPCKRKVPVYVVTNFMVVCVMAATALISFWSLRDY 397
DB 423 VREGAFGLLSVILGFSSFLIEPCKRKVGARLVNAMSNTFVAFACMAGTAITSLSVNEY 482
QY 398 --HGVQDAITANASIKAVCLVLFALFVGLVPLAIVVFPVTAQLAATRGGGGLCTGV 455
DB 483 ITEG-IOHAIGENRAIKTASLVFALLGFLPISITVSPFSTAEITADTGGGGLAIGLV 541
QY 456 NISVIPVITIALGAGPMDALFGKGNIPAFGVASAFALVGVGVVFLFPKISKROFRAYS 515
DB 542 NLAIVIPVITIALGAGPMDALFGKGNIPAFVLAALFALAAGVIATLKLPLNLSSSSYK--S 599
QY 516 AGGH 519
DB 600 SGFH 603
```

## ...RESULT 11

```
Q9FVL6 ID Q9FVL6 PRELIMINARY; PRT; 604 AA.
AC Q9FVL6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Sucrose transporter-like protein.
GN SUT2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20359188; PubMed=10899981;
RA Barker L., Kuhn C., Weise A., Schulz A., Gebhardt C., Hirner B.,
RA Hellmann H., Schulze W., Ward J.M., Frommer W.B.;
RT "SUT2, a putative sucrose sensor in sieve elements."
RL Plant Cell 12:1153-1164(2000).
DR EMBL: AF166498; AAG12987.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR TIGRFAMs: TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 604 AA; 65082 MW; 3403349B1692B2AF CRC64;
```

Query Match 56.7%; Score 1549.5; DB 10; Length 604;  
Best Local Similarity 54.8%; Pred. No. 1.1e-111;  
Matches 298; Conservative 73; Mismatches 116; Indels 57; Gaps 5;

QY 30 SLGRLLIAGMAGVGVQGWALQSLTTPYVQTGLSHALTFSFMLCGPIAGLVQPLVGL 89

Db 62 SLTLILSCTVAAGVQFGMALQSLTLPYIQTGLGIEHAFSSFIWLCGPITGLVVPQCVGI 121  
QY 90 YSDCTSRGRRRPFIITGCMILCAVAVIVGFSDDIGAALGDTKEHSLYHGPWHAAIV 149  
Db 122 WSDCKSKYGRRRPFIIGAVMISIAVIIIGFSADIGLGDTHKESCTFKGTRRAIV 181  
QY 150 YVGLFWLLDSNNITVQGPAMMADLCHDHGSPSAANSIFCSWALGNILGYSGSTNNWH 209  
Db 182 FVVGFWMLDLANNITVQGPAMMADLCHDHGSPSAANSIFCSWALGNILGYSGSTNNWH 241  
QY 210 KWPFPLKTSACCEACANLKGAFLVAVVFLVCLTVTLIFAKEVP----- 253  
Db 242 RWPFPLNRAACCEPCGNLKAFLVAVVFLVCLTVTLIFAKEVP----- 301  
QY 254 -----YR-----ANENLPTTKAGGEVETPTGPL 277  
Db 302 DSPQNTGFDLSQSKRELQYNSVANNESEMHVADNSPKNEBPDKDQDGSFADSPGAV 361  
QY 278 AV--LKGFKDLPPGMPSVLLVTAITWLSWPFILYDTHMGREIYHGDGPKGNSNAQISAFN 335  
Db 362 LVNLLTSLRHLPPAMHSVLIVMALTWLSWPFILYDTHMGREIYHGDGPKGNSNAQISAFN 421  
QY 336 EGVVGAFAFGLLNSVILGFSSEFLIEPCKRVGVVWVTSNFWVCVMAATALISFWSLR 395  
Db 422 QGVREGAFGLLNSVILGFSSEFLIEPCKRVGVVWVTSNFWVCVMAATALISFWSLR 481  
QY 396 DYHGYVQDAITANASIKAVCLVLFARLGVPLAILYSVPFAVTAQLAATRGGGGGLCTGV 455  
Db 482 ANTOGVQHVIGATSTOIALVFLSGLIPLAVTYSVPFSTAITELTADAGGGGGLAIGVL 541  
QY 456 NISTVPIQVIALGAGPMDALFGKGNIPAFGVASAFALVGVGVGVFLPKIKRQFRAYS 515  
Db 542 NLATVPMQVSVLCAAGPMDALFGKGNIPAFGVASAFALVGVGVGVFLPKIKRQFRAYS 598  
QY 516 AGGH 519  
Db 599 TGFH 602  
RESULT 12  
ID 080605 PRELIMINARY; PRT; 594 AA.  
AC 080605;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DE Putative sucrose/H+ symporter (Sucrose transporter).  
GN T17M13.3 OR SUC3.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
RA Shen M., Renning C.M., Fraser C.M., Somerville C.R., Venter J.C.,  
RT "Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence.",  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Meyer S., Truernit E., Huenner C., Besenbeck R., Stadler R., Sauer N.,  
RT "AtSUC3: an unusual sucrose transporter from Arabidopsis.",  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; AC004138; AAC32907.1; -  
DR EMBL; AJ289165; CAB92307.1; -  
DR InterPro: IPR003662; sub.transpporter.  
DR Pfam: PF00083; sugar.tr.1.  
DR TIGRFAMs; TIGR01301; GPH\_sucrose; 1.  
KW Transmembrane.  
SQ SEQUENCE 594 AA; 63972 MW; DF075A3601299DA0 CRC64;

Query Match 55.8%; Score 1523; DB 10; Length 594;  
Best Local Similarity 53.6%; Pred. No. 1.2e-109;  
Matches 288; Conservative 82; Mismatches 117; Indels 50; Gaps 5;  
QY 30 SLGRLLIAGVAGVQVGMALQSLTLPYIQTGLSHALTSMWLCGPIAGLVVQPLVGL 89  
Db 59 SLVTLVLSCTVAAGVQFGMALQSLTLPYIQTGLSHAFSSFIWLCGPITGLVVPQCVGI 118  
QY 90 YSDCTSRGRRRPFIITGCMILCAVAVIVGFSDDIGAALGDTKEHSLYHGPWHAAIV 149  
Db 119 WSDCKSKYGRRRPFIIGAVMISIAVIIIGFSADIGLGDTHKESCTFKGTRRAIV 178  
QY 150 YVGLFWLLDSNNITVQGPAMMADLCHDHGSPSAANSIFCSWALGNILGYSGSTNNWH 209  
Db 179 FVVGFWMLDLANNITVQGPAMMADLCHDHGSPSAANSIFCSWALGNILGYSGSTNNWH 238  
QY 210 KWPFPLKTSACCEACANLKGAFLVAVVFLVCLTVTLIFAKEVPYRANE----- 258  
Db 239 EWPFPLTSRACCAACGNLKAFLVAVVFLVCLTVTLIFAKEIPFTSNKPTRIQDSAPLL 298  
QY 259 -----NLPTTKAG-----EVETPTGPLAV---LKG 282  
Db 299 DLQSKGLEHSLKLNNGTANGIKYVERVDTDEQFGNSENEHQDETYVDGPGSVLNLTS 358  
QY 283 FKDLPPGMPSVLLVTAITWLSWPFILYDTHMGREIYHGDGPKGNSNAQISAFNEGVVGA 342  
Db 359 LRHLPPAMHSVLIVMALTWLSWPFILYDTHMGREIYHGDGPKGNSNAQISAFNEGVVGA 418  
QY 343 FGLLNSVILGFSSEFLIEPCKRVGVVWVTSNFWVCVMAATALISFWSLRDYHGVVQ 402  
Db 419 LGLLNSVILGFSSEFLIEPCKRVGVVWVTSNFWVCVMAATALISFWSLRDYHGVVQ 478  
QY 403 DAITANASIKAVCLVLFARLGVPLAILYSVPFAVTAQLAATRGGGGGLCTGVLSIVIP 462  
Db 479 YIMRGNETTAAVIVFALLGFLAITYSVPFSTAITELTADAGGGGGLAIGVLALVIP 538  
QY 463 QVIALGAGPMDALFGKGNIPAFGVASAFALVGVGVGVFLPKIKRQFRAYSAGGH 519  
Db 539 QMIVSLGAGPMDQLFGGGLNLPFAVLASVAAFAAGVIALQRLPTLSS-SFK--STGFH 592  
RESULT 13  
Q857X2  
ID 0857X2 PRELIMINARY; PRT; 297 AA.  
AC 0857X2;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DE Sucrose transporter, 5'-partial (Fragment).  
GN OSJNBA0091P11.1.  
OS Oryza sativa (Japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPONBARE;  
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,  
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,  
RA Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pal G.,  
RA VanAken S.E., Ulterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,  
RA Salzberg S.L., White O., Fraser C.M.;  
RT "Oryza sativa chromosome 3 BAC OSJNBA0091P11 genomic sequence.",  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC073556; AAL84308.1; -  
FT NON\_TER 1  
SQ SEQUENCE 297 AA; 31370 MW; BD332D42C395F9D7 CRC64;

Query Match. 48.1%; Score 1312.5; DB 10; Length 297;  
Best Local Similarity 83.4%; Pred. No. 1e-93;  
Matches 247; Conservative 25; Mismatches 23; Indels 1; Gaps 1;

```
QY 223 ACANLKGAFVAVVFLVLCITVTLIFAKEVPYRANENLPTTKAGGEVETETGPLAVLKG 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 ACANLKGAFVAVVFLVLCITVTLIFAKEVPYRANENLPTTKAGGEVETETGPLAVLKG 59
QY 283 FKDLPPGMPVLLVTAITWLSWFFPILYDWMGREIYHGDGPKGNAQISAFNEGVRVGA 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 FRLNLTGMPVLLVTAITWLSWFFPILYDWMGREIYHGDGPKGNAQISAFNEGVRVGA 119
QY 343 FGLLNSVILGSSFLIEPCMKRKGPRVWVTSNFMVCMVAMAATALISFWSLRDHYGYVO 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 FGLLNSVILGSSFLIEPCMKRKGPRVWVTSNFMVCMVAMAATALISFWSLRDHYGYVO 179
QY 403 DAITANASIKAVCLVFLAFGLVPLAFLVSVPPFAVTAQAATRGCGGLCTGVLNLSIYIP 462
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 KAITADKSIKAVCLVFLAFGLVPLAFLVSVPPFAVTAQAATRGCGGLCTGVLNLSIYIP 239
QY 463 QVVIAGAGPMDALGKGNIPAFGVSAPALVGVGVFLLPKISKROFRAVSAGG 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 QVVIAGAGPMDALGKGNIPAFGVSAPALVGVGVFLLPKISKROFRAVSAGG 595

RESULT 14
Q41152 PRELIMINARY; PRT; 533 AA.
ID Q41152;
AC Q41152;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sucrose carrier.
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CARMENCITA; TISSUE=COTYLEDON;
RA Weig A., Komor E.;
RT "An active sucrose carrier (Srl1) that is predominantly expressed in
the seedling of Ricinus communis L.";
RL J. Plant Physiol. 147:685-690(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Z31561; CAA83436.1; -
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 533 AA; 56359 MW; 0643E25D525AAAD8 CRC64;

Query Match 41.9%; Score 1144; DB 10; Length 533;
Best Local Similarity 45.0%; Pred. No. 2.2e-80;
Matches 226; Conservative 99; Mismatches 151; Indels 26; Gaps 8;

QY 19 AAADVHVAPISLRLILAGVAGVQVQWALQSLTPYVOTLGLSHALTSMFMLCGPI 78
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 AGAAEPNSSP--LRKVMVASTAGIQFGWALQSLTPYVQVLLGIPHTWAFAIWLGP 82
QY 79 AGLVQPLVGLYSDRCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCSL 138
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 SGMVLQPIVGVHSDRCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCSL 137
QY 139 YHGPWRHAAIVVVLGFWLLDFSNNTVOGPARAMMADLCDDHGPS-----AANSIFCSNMA 193
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 DSKPPTRAIAIFVGVFWLLDVANNMLQGPCRALLAD---SGTSOKKTRTANLSTFFMA 194
QY 194 LGNILGYSSGSTNNHKKFPFLTKTSACCEACANLKGAFVAVVFLVLCITVTLIFAKEVP 253
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 195 VGNVLGYAAGAVTHLYKLPFPFTKTACDVCYCANLKSCEFISIVLLSLTVLALSIVKKEP 254
QY 254 Y---RANENLPTTKAGGEVETETGPL--AVLKGFKDLPPGMPVLLVTAITWLSWFFP 308
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 255 WSPQADVNAEDDTASQASSAQPMPEGEILGAFKNLKRPMWILLVTCNLIAWFFPEL 314
QY 309 LYDTDMGREIYHGDGPKGNAQISAFNEGVRVAGFGLLNSVILGSSFLIEPCMKRKY-G 367
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 315 LFDTDMGREIYHGDGSSGSAEQKLKLYDRGVAGALGMLNSVILGFTSLGVEVLARGVGG 374
QY 368 PRVWVTSNFMVCMVAMAATALI-----SFWSLRDHYGYVQDAITANASIKAVCLVFLAF 422
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 375 VKRLWGINVFLAVCLAMTVLTKQAESTRFATVSGGAKVPLPPPSGVKAGALFAVM 434
QY 423 GVPALILSVFPAVTAQAATRGCGGLCTGVLNLSIYIPQVITIALGAGPMDALGKGN 482
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 435 GVPQAITYSIFPALASIFESNTSGAGQGLSLGVLNLSIYIPQIVSVAAGPMDALGEGNL 494
QY 483 PAFGVASAFALVGVGVFLLP 504
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 495 PAFVGVAAALASGIFALTMLP 516

RESULT 15
O65929 PRELIMINARY; PRT; 501 AA.
ID O65929;
AC O65929;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sucrose/H+ symporter (Sucrose/proton symporter).
GN SUN1B OR SUN1A.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANTALISE; TISSUE=LEAVES;
RA Shakya R., Sturm A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANTALISE;
RA Sturm A.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANTALISE;
RA Shakya R.;
RL Thesis (2000), Department of Department of Botany,
University of Basel, Basel, Switzerland.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Y16767; CAA76368.1; -
DR EMBL; Y16766; CAA76367.1; -
DR EMBL; AJ303198; CAC19688.1; -
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 501 AA; 54066 MW; 2BDD6692C39ACA4 CRC64;

Query Match 41.2%; Score 1126; DB 10; Length 501;
Best Local Similarity 44.6%; Pred. No. 4.9e-79;
Matches 225; Conservative 102; Mismatches 142; Indels 36; Gaps 8;

QY 18 GAAAVVDHVAPISLRLILAGVAGVQVQWALQSLTPYVOTLGLSHALTSMFMLCGP 77
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 GATAAPPSPRSVLSRLILRVASVACGIQFGWALQSLTPYVQELGIPHTWASSIWLGP 73
QY 78 TAGLVQPLVGLYSDRCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCS 137
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 LSGLLVQPIVGVHSDRCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCS 133
QY 138 LYHGPWRHAAIVVVLGFWLLDFSNNTVOGPARAMMADLC--DHHGPSAANSIFCSNMA 195
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 134 M-----AIVAFVIGFWILDVANNMTQGPCRALLADLTGNDARRTRVANAYESLFAIG 186
QY 196 NILGSSGNTNNHKKWFFPKLTSCACCEACANLKGAFVAVVFLVCLVTTLIFAKEVPPYR 255
Db 187 NVLGATGAYSGHYKVFPSLTSCTINCANLKSAYIDIIFIITYISISAKERPRI 246
QY 256 ANENLPTTKAGGEVET---EPTGPLAVLKGFKDLPFGMPSVLLVTAITWLSWFFFLYDT 312
Db 247 SSQDGPQFSDGTAQSGHIEAEFLWELFGTFRLPGSVVILLVTLCLNWIWGFPEFLDFT 306
QY 313 DWNGREIYHCDPKGSNAQISAFNEGVRGAFGLLNSVILGFSFLIEPMCRKVGRPVW 372
Db 307 DWNGREIYGEPN-----QQQSYSDGVRMGAFGLMNSVILGITSVLMEKLCRIWGSFWM 362
QY 373 VTSNFWCVMAAATALISFWSLR-DY-----HGYQDAITANASIKAVCLVFLAFGLVP 425
Db 363 GLSNILMTICFFAMLLITFAKNMDVGTNPPNGVISA-----LIIVAILGIP 411
QY 426 LAILYSVPFAVTAQLAATRGSGGGLCTGVNLISIVIPQVITIALGAGPMDALFGKGNIPAF 485
Db 412 LAITYSPVALVSTRIESLGLGQGLSGVNLAIIVPQVIVSLVSGSPWDQLFGGNSPAF 471
QY 486 GVASAFALGVGVGVFLL--PKISK 508
Db 472 VVAALSAFAGLIALAIRPRVDK 496

RESULT 16
Q9S725
ID Q9S725 PRELIMINARY; PRT; 512 AA.
AC Q9S725;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Sucrose transporter SUT2B.
GN SUT2B OR SUT2A.
OS Apium graveolens (Celery).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX NCBI_TaxID=4045;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MATURE LEAF;
RA Noiraud N., Lemoine R.;
RT "Sucrose transporters in celery.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF167416; AAD45391.1; -
DR EMBL; AF167415; AAD45390.1; -
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; tigr01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 512 AA; 54426 MW; 2637535216FFLED2 CRC64;

Query Match 41.1%; Score 1122.5; DB 10; Length 512;
Best Local Similarity 46.3%; Pred. No. 9.5e-79;
Matches 228; Conservative 99; Mismatches 134; Indels 31; Gaps 11;

QY 28 PISGLRILAGVAGGVQVQWALQSLTTPYVQTLGLSHALTSPMNLGCGTAGLVQPLV 87
Db 28 PIATWKLILVAATAAGVQFGWALQSLTTPYVQTLGLPHKWAAYIWLCPISGLMVLQPIV 87
QY 88 GLYSDRCTSRWRRRPILTGCMLICVAVIVVGFSSDIGAALGD--TKEHCSLYHGRPH 145
Db 88 GYSDRCQSGFRRRPFIASGACVAISVILIGFAADIGYKAGDDMTKT-----LKPR-- 140
QY 146 AATVYLVGLFWLLDSNNVTQGPARAMMADLC--DHHGPSAANSIFCSMMALGNILGYSSG 203
Db 141 AVTGFVIGFWILDVANNMLOGPCRALLADLTGNDTRMRSANAFYSPFMAVGNILGYAAG 200
QY 204 STNNHKKWFFPKLTSCACCEACANLKGAFVAVVFLVCLVTTLIFAKE---VPIRANENL 260
```

```
Db 201 SYNLYKLFPSKTHACDLYCANLKSFIISVLLIEITVLALTVVREKQWSPDEADEEP 260
QY 261 PTKAGGEVET--EPTGPLAVLKGFKDLPFGMPSVLLVTAITWLSWFFFLYDTDNWGRE 318
Db 261 PSS---GKIPVFGELFGAL-----KDLPRMLMLLVVTCNLNIAWFPFIFLFDTNWGRE 311
QY 319 IYHGDPKGSNAQISAFNEGVRGAFGLLNSVILGFSFLIEPMCRKV-GPRVVVWTSNF 377
Db 312 IY---GGTAGQKLYDQGVYRGSGLLNSVILGTSIAVEYLVRGVGVKILWGLVNF 367
QY 378 MCVMAAATALISFWSLRD-YHYQVQDAITANASIKAVCLVFLAFGLVPLAILYSVPFAV 436
Db 368 LLAIGLVMTVWVSKVAQHQHGGANGQLPPSAGVKAGALSLSILGILPLSIFSIFFAL 427
QY 437 TQALATRGSGGGLCTGVNLISIVIPQVITIALGAGPMDALFGKNIIPAFGVASAFALVG 496
Db 428 ASYSGSGAGQGLSGVNLAIIVPQVIVSLVAGPFDLSLFGGGLNLPFVVGATSAISG 487
QY 497 VVGVLPLPKISK 508
Db 488 VLAIVLLPKPSK 499

RESULT 17
Q9FE59
ID Q9FE59 PRELIMINARY; PRT; 510 AA.
AC Q9FE59;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Sucrose transporter SUT4.
GN SUT4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsals.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG ERECTA;
RX MEDLINE=20407148; PubMed=10948254;
RA Weise A., Barker L., Kuhn C., Lalonde S., Buschmann H., Frommer W.B.,
RA Ward J.M.;
RT "A new subfamily of sucrose transporters, SUT4, with low affinity/high
RT capacity localized in enucleate sieve elements of plants.";
RL Plant Cell 12:1345-1355(2000).
DR EMBL; AF175322; AAG09192.1; -
DR EMBL; AF175321; AAG09191.1; -
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; tigr01301; GPH_sucrose; 1.
SQ SEQUENCE 510 AA; 54835 MW; 986C68016D4CAE44 CRC64;

Query Match 41.1%; Score 1122; DB 10; Length 510;
Best Local Similarity 47.6%; Pred. No. 1e-78;
Matches 226; Conservative 85; Mismatches 144; Indels 20; Gaps 8;

QY 34 LILAGVAGGVQVQWALQSLTTPYVQTLGLSHALTSPMNLGCGTAGLVQPLVGLYSDR 93
Db 43 LLRVASVACGIGWALQSLTTPYVQELGIPHAWASVILWLCGSLGFLVQPLVGHSSDR 102
QY 94 CTSRWRRRPILTGCMLICVAVIVVGFSSDIGAALGDTEHCSLYHGRPHAAIVVVLG 153
Db 103 CTSKYRRRPFIAGAVAISVIVVIGHAADIGHAFDREGKIK----PR--AIVAFVLG 156
QY 154 FWLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSMMALGNILGYSSGNTNNHKKW 211
Db 157 FWLIDVANNMTQGPCRALLADLTENDNRTRRVANGYFSLPMAVGNVLGYATGYSNGWYKI 216
QY 212 FPELKTSCACCEACANLKGAFVAVVFLVCLVTTLIFAKEVPIRANENLPTTKAGGEVET 271
Db 217 FTETKTACVNECANLKSFAFIDVFTAITTILSVSAHEVPLAS----LTSEAHGQTS 272
```

```

QY 272 EPTGPIAVLKG-FKDLPPGMPSPVLLITATITWISWEPFIILYDTDMWGRIYHGDPKGSNAQ 330
Db 273 TDBAFUSEIFGTGTRYPPGNWIIITALTATGTGWEPFIILDTDMWGRIYHGDPEPNIG--- 329
QY 331 ISAFNEGVRVGFAGLLLSVILGFSFLLIEPCKRKVGPRVWTSNFMVVCVMAAATALIS 390
Db 330 -TYSAGVSMGALGLMNSVFLGITSVLMEKLCRWKGAFVWGISNLMIAICFLGLMITS 388
QY 391 FWSLRDHYGVQDQAITANASIKAVCLVLVFAFLGVPLAIIYSVPFAVTAQLAATRGGGQGL 450
Db 389 F--VASHLGVYIGHE-QPPASIVFAAIVLFIETILGIIPLAITYSPYALISIRIESLGIQGL 445
QY 451 CTGVLNLSIVPOVITATGAGPMDALFNGKNTIAPFGVASAFALVGVGVGVFLPK 505
Db 446 SLGVNLNLAIVPOIVSVSGPPDQLFGGGNSPALAVGAATGFIGGVAVAILAPR 500

RESULT 18
Q8VYX4:
ID Q8VYX4 PRELIMINARY; PRT; 513 AA.
DC Q8VYX4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Q8VYX4: sucrose transporter SUC1.
OS Suc1.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eucoids II; Brassicales; Brassicaceae; Brassica.
OC NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RP Coupe S.A., Sinclair B.K., Bucknell T.T., Greer L.A., Eason J.R.,
RA Heyes J.A.;
RT "The isolation and characterization of sucrose transporter homologs
RT from broccoli and their role in sucrose mobilization during
RT senescence";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY065839; AAL58071.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
SQ SEQUENCE 513 AA, 54663 MW: 0279B262E2291B2B CRC64:

```

Qy	363	CRKV-GRPVVWVTSNFWCVAMAAATALISTFWSLRDTHGYVDQDAITANASIKAVCLVLP	PAF 421
		:     :     :     :     :     :     :	
Db	355	GKKYGGAKRLWGIWVNFLLAIGLAMTVLVTKLA-ADYRKVAGPYAGSPGIRAGALS	SLFAV 413
Qy	422	LGVPLATILYSVPFAVTAQLAATRGGGGCLTCGVLTNTSIVIPQVITIALGAGP	WDALFGKN 481
Db	414	LGIPLATFTSPALASIFSSSSGAGGGLSLGVNLNLAIQVIMVSLGGGPPD	ALFGGN 473
Qy	482	IPAFGVASAFALGVGVGVFLPP	504
Db	474	LPAFIVGAIAAAISGVLAITVLP	496
RESULT 19			
Q9ZTB9	Q9ZTB9	PRELIMINARY;	PRT; 512 AA.
ID	Q9ZTB9		
IC	Q9ZTB9;		
DT	01-MAY-1999	(T'EMBLrel. 10, Created)	
DT	01-MAY-1999	(T'EMBLrel. 10, Last sequence update)	
DT	01-JUN-2002	(T'EMBLrel. 21, Last annotation update)	
DE	Sucrose transporter.		
GN	SUT1.		
OS	Apium graveolens (Celery).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; euasterids II; Apiales; Apiaceae; Apium.		
CC	NCBI_TaxID=4045;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	TISSUE=LEAF;		
RA	Noiraud N., Delrot S., Lemoine R.;		
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
DR	EMBL: AF063400; AAC99332.1; -.		
DR	InterPro: IPR003562; sub_transporter.		
DR	Pfam: PF00083; sugar_1.		
DR	TIGR:TRAMS; TIGR01301; GPH_sucrose; 1.		
KW	Transmembrane.		
SO	SEQUENCE 512 AA: 54520 MW: 4D76A7854ADF903		CRG64;

```

Db 431 YSSGAGAGLGLSLVNLAIWVPMIVSLVGLPPDLSLFGGNNLPFAVYVGAISAISGVLA 490
Qy 500 VELLPKISK 508
Db 491 IVLLPKRCK 499

RESULT 20
Q9M3R4 PRELIMINARY; PRT; 510 AA.
AC Q9M3R4
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sucrose transporter (Putative sucrose/H+ symporter).
GN SUC4 OR ATIG09960.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RN NCBI_TaxID=3702;
RP SEQUENCE FROM N.A.
RA Sauer N.K.;
RT "An Arabidopsis thaliana sucrose transporter.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2].
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Karlini P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RA "Full Length cDNA of gene Atlg09960 (GI:15218362).";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ289166; CAB92308.1;
DR EMBL: AV072092; AAL59915.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRfams: TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 510 AA; 54784 MW; 3AF99CF2D48ACBD4 CRC64;

Query Match 40.9%; Score 1117; DB 10; Length 510;
Best Local Similarity 47.4%; Pred. No. 2.5e-78;
Matches 225; Conservative 85; Mismatches 145; Indels 20; Gaps 8;

Qy 34 LILAGMVAGGVQVQGWALQSLTPYVQTLGLSHALTSMFLCGPIAGLVQPLVGLYSR 93
Db 43 LLRVASVACGIFGQWALQSLTPYVQTLGLSHALTSMFLCGPIAGLVQPLVGLHSR 102
Qy 94 CTSRWGRRRPFILTCMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVVVLG 153
Db 103 CTSKYGRRRPFIVAGAVAISVNVIGHAADIAGWAFDREGKIK----PR--AIVAEVLG 156
Qy 154 FWLLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSHMLGNILYSGSGSTNNHKK 211
Db 157 FWLLDVANNNTQGPCRALLADLTENDNRRVANGVYSLEFMAVGNVLGYATGNGWYKI 216
Qy 212 FFLKTSACCEACANLKGAVLVAVVFLVCLTFLIPAKEVPYRANENLPTTRAGGVEV 271
Db 217 FTFKTVACNVCANLKSAYIDVFTAITILSVSAHEVPLAS-----LASEAGTSG 272
Qy 272 EPTGPLAVLKG-FKDLPPGMPSSLVTAITWLSWFFPILYDTDMGREGIYHGDPKGNAQ 330
Db 273 TDEAFLEIFGTRYFPNGWIIILLVTAIWIGWFFPILYDTDMGREGIYGGEPNIG--- 329
Qy 331 ISAFNEGVVGAAGLLNSVILGFSSFLIEPMCKRKGVPVWVTSNFMVCMVMAATALIS 390
Db 330 -TYSAGVSMGALGLMLNSVFLGITSVLMKLCRKGAGFVWGISNLMIAICFLGMITTS 388

```

```

Qy 391 FWSLRDHYGVQDDAITANASTKAVCLVLEAFVLGVPLATLYSVFAVTAQLAATRGGOGL 450
Db 389 F--VASHLGVIGHE-QPPASIVFAAVLITFTLLGILAITISVPYALISIRIESLUGOGL 445
Qy 451 CTGVLNISVIVPQVIAIGALGPMDALFCKGNIPAGFVASAFALVGVGVVFLPK 505
Db 446 SLGVLNLAIVIPQVIVSGSGPMDQLFGGNSPALAVGAATGFIGGIVAILALPR 500

RESULT 21
Q40938 PRELIMINARY; PRT; 510 AA.
AC Q40938
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sucrose transporter.
GN PTP1.
OS Plantago major (common plantain).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veroniceae; Plantago.
RN NCBI_TaxID=29818;
RP SEQUENCE FROM N.A.
RA MEDLINE=95093474; PubMed=8000426;
RA Gahriz M., Stolz J., Sauer N.;
RT "A phloem-specific sucrose-H+ symporter from plantago major L.
RL Plant J. 6:697-706(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: X75764; CAA53390.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRfams: TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 510 AA; 54018 MW; C6905E27F9A80B85 CRC64;

Query Match 40.9%; Score 1116; DB 10; Length 510;
Best Local Similarity 44.0%; Pred. No. 3e-78;
Matches 227; Conservative 103; Mismatches 158; Indels 28; Gaps 11;

Qy 18 GAAAVV-----DHVAPISLRLILAGVAGGVQVQGWALQSLTPYVQTLGLSHALTSE 71
Db 9 GAKTIITTPPEEAPIR--NIFLVAIAAGVQFGWALQSLTPYVQTLGLGPKHWASY 66
Qy 72 MWLCGPIAGLVQPLVGLYSRDRCTSRWGRRRPFLITGCMILICVAVIVGVFSSDGAALGD 131
Db 67 IWLCPISGIMVQPVGVFSDNCTSRFGRRRPFIAAGLVGVAVVLIGFAADLGHAGGD 126
Qy 132 TKHCISLYHGPRHAAIVVVLGVFLDPSNNTVQGPARAMMADLCDDHGP--SAANSTFC 189
Db 127 -----SLGDGLKPRAGVYFVFGFWILDVANNMLOGPCRALLADLGGNTKKMANAFFS 181
Qy 190 SMWALGNILYSGSGSTNNHKKWFFPKTSACCEACANLKGAVLVAVVFLVCLTFLVIFA 249
Db 182 FEMAVGNVLGYAAGSIRMYKVRPFESKTKACDIYCANLKSFIITLITLTTLALSIV 241
Qy 250 KEVPIRANENLPTTKAGEVETEPTGLAVLKGFKDLPPGMPSSLVTAITWLSWFFPIL 309
Db 242 REKRHVAEQVTAARKGFKIPVFP-----ELFGALKDLPMPMWVLLVTALNMIAWFGFL 297
Qy 310 YDTDMGREGIYHGDPKGNSQAISA--FNEGVRVCAFGLLNSVILGFSSFLIEPMCKRV- 366
Db 298 FDTDMGREGY-GETQHKAPELAVIYKGVSAAGALMLNSIVLGFSAAGLVQVWMAALG 356
Qy 367 GPRVWVTSNFMVCMVMAATALISFWSLRDHYGVVQDAI--TANASIKAVCLVLEAFVLG 424
Db 357 GVKRLGWVNFILAIICLMTIVIT--KVASHRYPYSGVLOTPESSVKIGALVVFSAIGI 414
Qy 425 PLATLYSVFAVTAQLAATRGGOGLCTGVNLINISIVIPQVIAIGALGPMDALFCKGNIP 484

```

Db 415 PLATFSPVPALASIYSTTTGSCQGLSLGLVNLNLAIVIPQMVSVASGPPDWMFGGGNLPA 478

Qy 485 FGVAFAALYGVGVGVVFLLPKIKSKROFRAVS-AGGH 519

Db 475 FVGVAAAAAGSIFAFTMLPSPAESIKNLSVAGGH 510

RESULT 22

QX9HL6 ID Q9XHL6 PRELIMINARY; PRT; 524 AA.

AC ACQ9XHL6;

DT 01-NOV-1999 (T-EMBLrel. 12, Created)

DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)

DE Sucrose transport protein SUT1.

OS Pisum sativum (Garden pea).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.

OX NCBI\_Taxid=3886;

RN [1]

RX SEQUENCE FROM N.A.

RX MEDLINE=99291546; PubMed=10363367;

RA Tegeder M., Wang X.D., Frommer W.B., Offler C.E., Patrick J.W.;

RA "Sucrose transport into developing seeds of *Pisum sativum* L.;"

PL Plant J. 18:151-161(1999).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY);

DR EMBL; AF109922; AAD41024.1; "

DR InterPro; IPR003662; sub\_transporter.

DR Pfam; PF00083; sugar\_tr; 1.

DR TIGRfams; TIGR01301; GPH\_sucrose; 1.

KW Transmembrane.

QX SEQUENCE 524 AA; 55279 MW; B5CE3F880D2C1E6B CRC64;

Query Match 40.9%; Score 1116; DB 10; Length 524;

Best Local Similarity 44.2%; Pred. No. 3.1e-78;

Matches 221; Conservative 106; Mismatches 153; Indels 20; Gaps

Qy 28 PISLGRILLAGVAGVQYGVWALQLSLTTPYVOTGLSHALTSFMMLCGPIAGLVVQPIV 87

Db 32 PSPLRKIMVVASIAAGVQFGWALQLSLTTPYVOLLGHHTWAAIWLCPISGMLVQPVV 91

Qy 88 GLYSDRCTSRWGRRRPILTCMLICVAVIVGVSSDIGAALGDTKEHSLYHGPWKHA 147

Db 92 GYHSDRCTSREGRRRPFAAGSTAVAIAPVLIGYAADLGHSGFD----NLQKKVRPRAI 146

Qy 148 IVYVLGFWLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSNMALGNILGYSSGT 205

Db 147 GIFVVGWILDVANNMLQGGPCRALLDGDCAGNQRKTNRNANAFSFMVGNVLGYAAGY 206

Qy 206 NNHWKHPFPFKTSACCACANLKGFLVAVVVLCLTTLTFAKEVPYRANENLPTTKA 265

Db 207 SKLYHVFPFKTACNVYCANLKSCEFLSIALTLVLTATLIVYKETPLIAEKAVVTAED 266

Qy 266 GGEVETEPT-GPLAVLKGFDLPPGMPSVLLVTAITWLSWFFPILYDTDMWGREIYHGP 324

Db 267 GSGNGMPCFQLS--GAPKELRPMWILLVTLCLNIAWAFPELLEFDTDMGKEVY--- 320

Qy 325 KGSNAQISARNEGVRVGAFLNLSVLGFSFLIEPMCRKV-GPRVVVWTSNFWVCVAM 383

Db 321 GGTVGEGHAYDMGVRAGLGLMLNSVVLGATSLGVDIRLARGVGVKRLMGIVNFLAICL 380

Qy 384 AATALISFWS--LRDY--HGYYQDAITANASTKAVCLVFLAFLGVPLAILYSVPPAVTA 438

Db 381 GLTVLVTKLAQHSRQVAPGTGGQLDPLPSPGGIKAGALTFLSVLGIPLAITIYSPALAS 440

Qy 439 QLAATRGGCGGLCTGVNLNISIVPIQVITIAFLKAGPMDALFCGNIPAFGVASAFALYGVV 498

Db 441 IFSSTSGAGGLSIGVNLNLAIVIPQMVSVLSPGMDALFGGNLPAFVVGVAVALASGLI 500

Qy 499 GVFLLPKIKSKROFRAVSAGG 518

Db 501 SMILLPSPPPDMAKSVSATG 520

```

RESULT 23
Q9SP63 PRELIMINARY; PRT; 501 AA.
AC Q9SP63;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Sucrose transporter.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC Vitis.
OX NCBI_TaxID=29760;
RN [1]
SEQUENCE FROM N.A.
RP STRAINS=CV. UGNI BLANC; TISSUE=GRAPE BERRY;
RA AGEORGES A., Issaly N., Picaud S., Delrot S., Romieu C.;
RT "Identification and functional expression in yeast of a grape berry
RT sucrose carrier."
RL Plant Physiol. Biochem. 38:177-185(2000).
CC -1. SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF182445; AAD55269.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR01301; GPh_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 501 AA; 54075 MW; 4FD19DFAFES39077 CRC64;

Query Match 40.8%; Score 1115.5; DB 10; Length 501;
Best Local Similarity 45.6%; Pred. No. 3.2e-78;
Matches 232; Conservative 90; Mismatches 152; Indels 35; Gaps 10;

QY 17 RGAADVHVAP-ITGLRLILAGWAGVGYGNALQLSLTTPVYQTLGLSHALTSPMLC 75
DB 12 RPRALICEPVPRVPLRRLLRVASVACCIQFGNALQSLTTPVYQELGIPHANSSIILC 71

QY 76 GPTAGLVQPLGLYSDRCTSRMGRRRPFILTGOMLCIVAVIVVGSSDITGAALGDTKEH 135
DB 72 GPLSGLLVQPLVGHLSDRCSNRCFRGRRPFIVAGATSIWVAVLIGFSTDTGGLLGDCADR 131

QY 136 CSLYHGPRWHAIVYVLGFWLLDFSNNTVOGPARAMADLC--DHHGPSAANSIFCSMA 193
DB 132 -----RPRAVATVVGFWLLDVANNVTGQPCRALLDATKDHRRTRVNAVYFSLFIA 184

QY 194 LGNITLYSSGSTNNHKKWFFPKLTACCACANLKAFLVAVVLCITVTLIFAKEVP 253
DB 185 VGNVLGFATGYSQWPRIFWFTSTSSCNADCNLKS AFLDIIIFAIITTVISITAAQELP 244

QY 254 -----YRANEMLPYTKAGGEVETPTGPLAVLKGFKDLPGMGSPVLLVTAITWLSWF 305
DB 245 LSSSSRSTHISEMAESTHAQEAFLWELFTLRLYSG-----SIWILFVTALTWIGWF 298

QY 306 PFILYDPTDMWGREIYHGDG-KGSNAQISAFNEGVRVYGAFLLLNSVILGFSFSLIEPMCR 364
DB 299 PFLLDPTDMWGREIYHGKPNQGN-----YNTGVRMGALGLMLNSVVLGITSVLMKLCR 353

QY 365 KVGPRVVVYVTSNFMVCVYMAATALIS-FWSLRDHYGVQDAITANASIKAVCLVLFAPFLG 423
DB 354 KWGAGFWVGLSNILMSCFLMLILSAVVKHMDFLGH--DLPPSGVVVIAA--LIVESILG 409

QY 424 VPLAIIYVFPVNTAQLAATRGGGGLCTGVNLISVIPQVITAGLACGPDALFGKGNIP 493
DB 410 IPLAITYSVPYALISTRIESLGLGQGLSMGVNLAIVIPQVIVSLGSGPMDQLFGGNSP 469

QY 484 AFGVAFAFALVGGVGVFLPLPKISKROFR 512
DB 470 SLAVAAYAAFAAGLVAITAIIPRSADKSR 498

RESULT 24
Q9SLN7

```



[illegible]

RESULT 25	
065803	
ID	065803
AC	065803
DT	01-APR-2008
DT	01-APR-2008

